

A typological approach to West Polesian Morphology and Syntax

by

Kristian Roncero Toledo



Submitted for the Degree of Doctor of Philosophy

University of Surrey

Faculty of Arts and Social Sciences

School of Literature and Languages

Supervisors:

Dr. Matthew Baerman

Prof. Greville G. Corbett

Declaration of originality

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above,

Kristian Roncero,

University of Surrey, Guildford

23rd of January 2019

Summary

West Polesian is an understudied Eastern Slavonic variety spoken between Belarus, Ukraine and Poland. The speech community has been geographically isolated for centuries, contributing to the preservation of older forms of Slavonic, but also the development of innovations. Nowadays it is surrounded by several closely related and standardised varieties that cause interference. This work provides a typological approach to the particularities that distinguish West Polesian morphology and syntax from its neighbours. Besides documenting and describing the morphology and syntax, I deal with theoretical questions that arise in the light of the data collected through fieldwork. I devote most of my attention to three phenomena. The first one is the adnumerative form, a morphosyntactic form that only appears when a noun is headed by specific numerals. I start by describing numerals and numeral phrases in West Polesian, and I narrow the focus to the interactions with nouns. The morphosyntactic nature of the adnumerative is complicated: it displays properties of both case and number values, but neither of them fully. Using insights from Canonical Typology I give an analysis of its morphology, syntax and the canonical behaviour of features and values. The second one is suppletion, which strongly characterises the adnumerative cells of the noun paradigms. I use the suppletive nouns ‘year’ and ‘person’ to see what they can teach us about suppletion in general. Still within the framework of Canonical Typology, I show how these nouns approach the canonical instance of suppletion and how they have the most complex paradigms in the Slavonic family. And the third one is the surpassingly large inventory of constructions West Polesian has developed to express futurity. I evaluate whether they are all legitimate (i.e. grammaticalised) constructions; and if so, what is their origin and relation to other future tense constructions in Europe.

Acknowledgements

There are many people without whom I would have not been able to do this work and to whom I am endlessly thankful. First of all, I would like to thank all the West Polesian speakers, this is also their project and they should be honoured for it (whilst I assume the responsibility for all the mistakes). They have been extremely patient with me, hospitable and warm. I feel enormously privileged by the opportunity of meeting such a nice people. I display the names of those participants who have consented to do so in (*Appendix I*), but I would like to specially thank my two main informants Baba Źenja from Torokan (who unfortunately is no longer with us), Maryja Handryjŭna from Bodanywka, and her son Viktor Aljakseevič, who arranged almost half of the interviews for me and took me around the region and always made sure I was comfortable. I would also like to thank the Omelyančuk family, for being my second family in Belarus.

The fieldtrips where I obtained the data have been possible thanks to the generosity of The Gilchrist Educational Trust (UK), The Philological Society (UK) and the Belarusian Charity (UK), as well as the University of Surrey FASS Fellowship which covered my tuition fees and living expenses. Special thanks to Steven Marlett, José Herrero and Iván Igartua, for all their support and advice, especially at the very initial stage of the project and for their help in applying for the scholarship.

Many academics have taken care of my research and provided me useful material. Among these, particularly thanks to Nina Barszczewska, Mirosław Jankowiak, Yuni Kim, Lena V. Levancèvič, Nerea Madariaga, Lidia F. Mazzitelli, Erich Round, Nicholas

Sims-Williams and Anna Thornton. But particularly thank you to Fëdar D. Klimčuk and Ivan U. Sacuta for introducing me to the world of West Polesian.

Thank you as well to all the people at the Anglo-Belarusian Society and the Belarusian Library in London, especially Ihar and Fr. Siarhey.

I do not have words to thank express my gratitude for all the support and advice I have obtained from my supervisors, Matthew Baerman and Greville Corbett. But also my colleagues and visitors at SMG. To Oliver Bond, for being like an older brother in SMG and academia. To Borja, Helen, Jérémy and Mike for their help reviewing, to all the SMG colleagues (in particular to the “Fridayclub”) for all the fun, good advice and patience with me.

And finally, I could have never finished this without the support of my amazing parents and friends. Particularly thank you to Andrew, Dave, David, Laura, Lina, Martin, Heather, James, Jim, Joseba, Jordan and Sasha for being with me in my darkest days. Not all angels have wings.

Contents

Declaration of originality	iii
Summary.....	iv
Acknowledgements.....	v
Part I Setting the scene	9
Chapter 1 Introduction.....	10
1.1. Western Polesie: geography, culture and language.....	10
1.2. Sociolinguistic overview.....	13
1.2.1. Autoethnonyms and autoglottonyms.....	14
1.2.2. History of Belarusian.....	17
1.2.3. History of West Polesian	18
1.3. Previous research	21
1.4. Writing conventions.....	25
1.4.1. List of abbreviations and glosses used in this work	29
Chapter 2 Methodology.....	31
2.1. Criteria for choosing the villages and the speakers.....	32
2.2. Data collection.....	34
2.2.1. Data-elicitation sessions.....	34
2.2.2. Classification of the data.....	38
2.2.3. Technical equipment.....	39

2.3. Limitations	39
2.3.1. Time, geographic and legal limitations	40
2.3.2. Funding	41
2.3.3. Standard language(s) bias	41
2.3.4. Gender imbalance	43
2.4. Ethical issues	45
2.4.1. Obtaining informed consent	45
2.4.2. Recognition of the contribution of the participants	46
2.4.3. Dissemination of results and archiving	46
2.4.4. Additional ethical measures to ensure human dignity	47
2.4.5. Remuneration for language consultants	48
2.5. Summary	49
Chapter 3 Typological overview of West Polesian	50
3.1. Phonology	51
3.1.1. Phonemic inventory	52
3.1.2. Palatalisation	55
3.1.3. Stress	57
3.1.4. Phonological and phonotactic rules in comparison with Eastern Slavonic	57
3.2. Morphology	63

3.2.1. Nominal Morphology.....	63
3.2.2. Verbal morphology.....	68
3.3. Syntax.....	71
3.4. ANIMACY.....	73
3.5. Summary	76
Part II. Main research outcomes	78
Section I. Numeral phrases and quantification	79
Chapter 4 Numeral phrases and numerals in West Polesian.....	80
4.1. General overview of numerals in West Polesian.....	81
4.2. Cardinal numerals.....	83
4.2.1. General morphological remarks.....	84
4.2.2. Morphosyntactic behaviour.....	89
4.2.3. Postnominal numerals.....	90
4.3. Collective numerals.....	91
4.3.1. Definition.....	92
4.3.2. Morphological and syntactic properties.....	93
4.3.3. Restrictions.....	98
4.3.4. Summary.....	111
4.4. Pronominal numerals	112
4.4.1. Definition.....	113

4.4.2. Properties	116
4.4.3. Cross-linguistic comparison.....	126
4.4.4. Summary.....	137
4.5. Other numerals.....	138
4.5.1. Ordinals	138
4.5.2. Fractions	140
4.5.3. Distributives	141
4.5.4. Quasi-adverbial numerals	143
4.6. Summary	145
Chapter 5 Adnumerative forms	146
5.1. Introduction.....	147
5.1.1. Methodological remarks	151
5.1.2. Cross-linguistic overview	152
5.2. Properties of the West Polesian adnumerative.....	162
5.2.1. Etymology.....	162
5.2.2. Morphology.....	167
5.3. Nouns headed by higher numerals.....	186
5.3.1. Introduction	186
5.3.2. Properties of the GREATER ADNUMERATIVE.....	193
5.3.3. Arguments against the existence of the GREATER ADNUMERATIVE	202

5.3.4. Conclusions	205
5.4. Syntactic analysis: perspectives.....	208
5.5. Typological analysis of the adnumerative forms	212
5.5.1. The ADNUMERATIVE as a NUMBER value.....	212
5.5.2. The ADNUMERATIVE as a CASE value.....	224
5.5.3. The West Polesian ADNUMERATIVE from the point of view of CT	234
5.5.4. Final hypotheses	243
5.6. Summary and conclusions.....	251
Chapter 6 Canonical suppletion in West Polesian: the nouns ‘year’ and ‘person’	255
6.1. Introduction.....	257
6.1.1. Presentation of the problem	258
6.1.2. Methodological remarks	260
6.1.3. Remarks about Slavonic counting systems and suppletion.....	261
6.2. The noun ‘year’.....	262
6.2.1. The noun ‘year’ across the Slavonic family.....	263
6.2.2. Etymology.....	265
6.2.3. West Polesian suppletion patterns for‘year’	266
6.2.4. Conditions. Are there any rules for the distribution of the stems?	269
6.3. The noun ‘person’	283

6.3.1. The noun ‘person’ across the Slavonic family	284
6.3.2. West Polesian suppletion patterns for ‘person’	287
6.3.3. Conditions for suppletion for the noun ‘person’	291
6.4. Challenges to the traditional assumptions about suppletion	302
6.5. The nouns ‘year’ and ‘person’ in the light of Canonical Typology	304
6.5.1. The noun ‘year’	305
6.5.2. The noun ‘person’	319
6.5.3. Overall summary	326
6.6. Summary and conclusions	328
Section II. Future tense	331
Chapter 7 West Polesian FUTURE TENSE	332
7.1. Introduction	333
7.1.1. The FUTURE in West Polesian	333
7.1.2. Theoretical problems about the FUTURE TENSE	338
7.1.3. Remarks about methodology	341
7.2. Uses and functions of the examined constructions	341
7.2.1. The SYNTHETIC FUTURE	342
7.2.2. DE-OBLIGATIVE FUTURE	348
7.2.3. The DE-VOLITIVE FUTURE	351
7.2.4. The DE-VENITIVE FUTURE	354

7.2.5. The PERFECTIVE FUTURE (<i>or</i> PRESENT)	359
7.2.6. The COPULAR FUTURE	361
7.2.7. Other marginal or ‘hapax’ constructions.....	363
7.3. Grammaticalisation.....	366
7.3.1. Definition.....	367
7.3.2. Grammaticalisation tests	367
7.3.3. Summary.....	397
7.4. West Polesian future-tense constructions from a historical-typological perspective	399
7.4.1. The SYNTHETIC FUTURE	400
7.4.2. The DE-OBLIGATIVE (or ‘have’) FUTURE	408
7.4.3. The DE-VOLITIVE FUTURE	422
7.4.4. The DE-VENTIVE future tense.....	430
7.4.5. PERFECTIVE FUTURE (or PRESENT).....	439
7.4.6. Slavonic COPULAR constructions	442
7.4.7. Hapax constructions	446
Part III. Summary and conclusions	451
Chapter 8 Summary and conclusions	452
References.....	457
Other resources referred to in this work	473

Appendix I. List of participants.....	475
Appendix II. Map of villages covered in this research	478

Part I Setting the scene

Chapter 1

Introduction

West Polesian (WP) is an Eastern Slavonic variety spoken in an area comprising the Polish region of Podlasie, the south-western half of the Brest region in Belarus, and the Volynsk region in Ukraine.¹ The speech community lives in an area of difficult access (as it is frequently flooded), which has contributed to the preservation of older stages of the Slavonic language and culture, but likewise to the creation of some innovations. That way West Polesian grammar shows unique features within the Eastern Slavonic subgroup. But, at the same time, being surrounded by other standardised and closely related languages, its grammar is changing under their pressure.

1.1. Western Polesie: geography, culture and language

In my MA thesis (Roncero 2015) I distinguished three concepts around Polesie and Polesian. The first one is the geographic territory of Polesie, which according to Vjarenič (2009) comprises the micro-region of Podlasie in Poland in the West, virtually all the areas around the Belarusian-Ukrainian border and the neighbouring region of Rjazansk (Russian Federation) in the East. The geographic area is characterised by lowlands and a marshy topography, which when the snow melts,

¹ There is a long debate on whether it is a language on its own or not. I have no interest on getting into that debate for the premises of this work, thus, I deliberately use the terms **SLAVONIC VARIETY** and **LANGUAGE** ambiguously in this thesis.

particularly during the springtime, causes frequent floods. The earliest references to Polesie come from the Greek historian Herodotus (Klimčuk 1992) who described the lowlands, and this is why Polesian marshes have been often referred to as “Herodotus’ Sea”. Starting in the seventies, but particularly during the next two decades, the Soviet government started a programme for draining and canalising the marshes (commonly known as *Meljoratsija* ‘improvement’) so that roads could be built to the villages. Nowadays most of the villages have roads which link them to larger population centres, although floods are still frequent in certain villages. As an anecdote, several speakers from Žydča (Pinsk district), who lived there before the *Meljoratsija*, shared memories of bread being delivered by helicopter and travelling in boats inside the village in the worst moments of the floods.

The second concept is the cultural or ethnic Polesie. This term is certainly more vague and harder to frame, but it involves the cultural practices of the ethnic *Poleščuki* (‘Polesians’). When it comes to physiognomy, it is popularly said (at least in Belarus) that they have a shorter height than the average (Belarusian) and darker hair, eyes and skin. There are other subtle details which distinguish them from the rest, such as the patterns they use for embroidery or the shapes and decorations in the roofs of their houses. The area of the cultural Polesie is certainly blurry and disputable, but it matches approximately the geographic extension of Polesie. However, the ethnic and linguistic differentiation of Polesians is a sensitive issue and I do not feel qualified to propose or defend any position. For this reason, I simply summarise some observations in the culture and mention some of the generally accepted facts about Polesians (I expand more on this in (§1.2.1.)).

Polesian culture has many elements in common with other Eastern Slavonic regions. Yet in many respects it retains elements of older stages of the Common Slavonic era. According to the ethnographic work of Tolstoj (1983) Polesie can be considered a model of the “Proto-Slavonic culture” in as much as it has been quite isolated from contact with other non-Slavonic groups/languages. The ethnographer Inna A. Šved (p.c.) also believes that together with the people in the Carpathian Mountains, Polesians are the Slavonic group with the most conservative culture. In addition to the aforementioned ethnographic research, I can provide evidence for treating West Polesian culture as conservative (without having to say that it is older than other groups). In the villages I have worked people have shared with me testimonies that reveal that magic and sorcery is very present in their lives (e.g. the existence of *šeptuxy* [ʃeptu'xɪ] ‘whisperers’ in virtually every village; or rituals to keep harmonious relationships with their dead relatives).

Finally, the third concept is the area of the linguistic extension of Polesian. Different dialectologists give slightly different classifications. For some there are West, Central and Eastern Polesian varieties (e.g. Del Gaudio 2014), whereas for others (e.g. Levancèvič 1994) the difference is only between West and East (at least inside Belarus). In any case, all agree that the varieties within West Polesian are the most distinct from the national standard varieties and the most special within the Polesian group. According to Fjodor D. Klimčuk (p.c.) the most significant (i.e. distinct and “pristine”) part of West Polesian is the area inside Belarus. In fact, West Polesian varieties are often referred to in the dialectological literature as “the Brest-Pinsk Polesian dialects” (e.g. Hulickaja et al. 1992), as the distribution of West Polesian

covers roughly the area below the imaginary line between Brest-Pinsk. Levancèvič (2013) gives the map of the varieties in the region of Brest (Belarus) in (Figure 1).

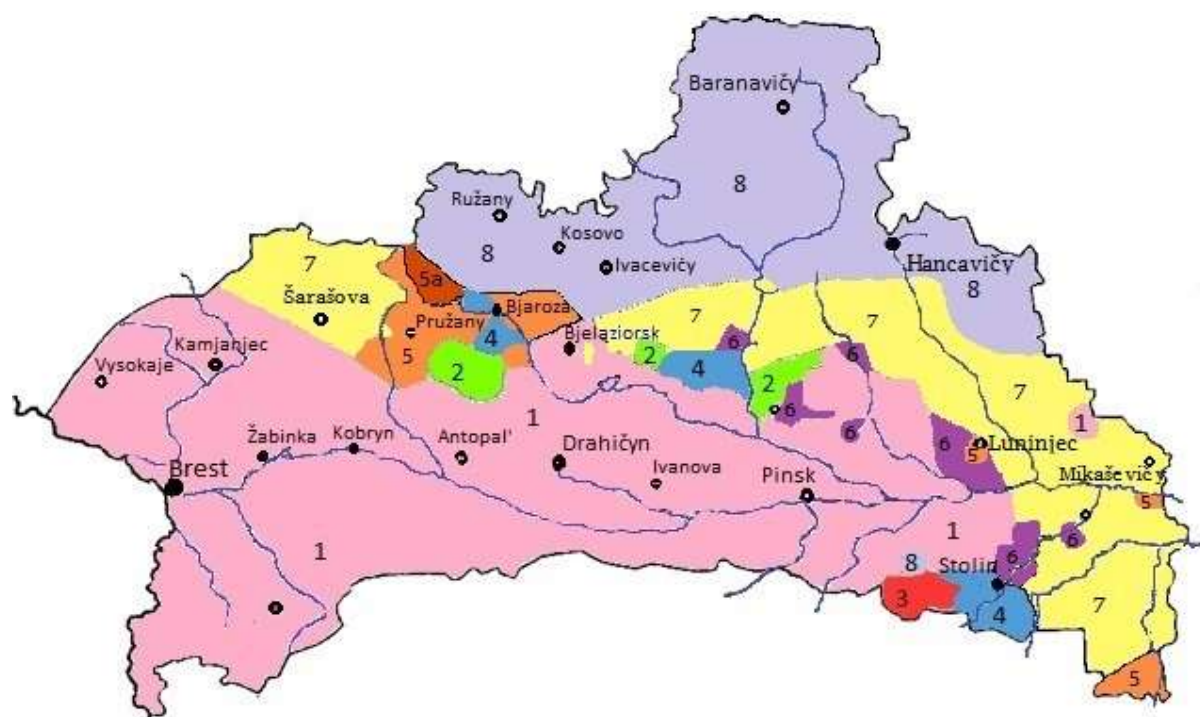


Figure 1 Map of the varieties in the region of Brest (Belarus), based on Levancèvič (2013) [My translation. Modified].

The area coloured pink (with the number 1) is the “proper” area of West Polesian and the focus of this research.

1.2. Sociolinguistic overview

There are no official statistics about the number of speakers. The UNESCO *Atlas of Endangered Languages* (2011 online version) estimated 600,000 speakers (probably only

in Belarus).² That number seems rather high if it only considers the speakers in Belarus. As I have already said, I undertook all my fieldwork in the region of Brest (Belarus). According to Belstat (2018), in the year 2018 the region of Brest had 1,384,500 inhabitants, and of those 29.49% (408,100) lived in rural areas. West Polesian is spoken mainly in rural areas, although as we can see from the map in (Figure 1) it only covers a bit less than a half of the region of Brest. Therefore, given the lack of official statistics, we could use this as a reference (i.e. circa 200,000 speakers in Belarus).

1.2.1. Autoethnonyms and autoglottonyms

I refer to this variety as West Polesian (also spelled “West Polissian”), as it was the most established term in the literature in English and (less) in Belarusian at the beginning of my research (Duličenko 1995, Gustavsson 1998). The term *Zaharoddzian* (from Belarusian *загароддзкія гаворкі*) has several centuries of tradition according to Klimčuk (1999), although he is probably the only one popularising it in the academic literature in Belarusian and Ukrainian. In any case, as far as I am aware, I am the only author who has ever made any mention of this term in the literature in English (Roncero 2016).

In my experience Polesians often lack a full “ethnic and/or linguistic self-awareness”, which as Dorian (2010: 287-288) puts, it is a widespread assumption in Western

² For reasons unknown to the author, the entry for Polesian was removed from the Atlas sometime between July and October 2015 (right at the beginning of the PhD research), but no explanation was given on the website.

societies and academia. Thus, this can often be confusing for the Western researcher.

Grace (1992) already brought up this issue after his work in Melanesia:

“One of the things I found most puzzling was that in some areas the people seem to have no conception of what their language is and no sense of belonging to a linguistic community” (Grace 1992: 122).

Most West Polesian speakers I have interviewed (in Belarus) were aware that they did not speak (Standard) Belarusian, (Standard) Ukrainian or Polish, yet many struggled to tell me the name of their language or their ethnonym. In Belarusian Polesie, where I conducted fieldwork, they refer to Belarusians as ‘they’, ‘them’, ‘those’, and very rarely *litvyny* [litvini].³

When it comes to the name of the language or variety, speakers and non-speaking locals (e.g. in the city of Brest) refer to it in many different ways, although some of those terms are more derogatory than others. The most common terms are [po dɪrɪ'vjenskomu] (lit.) ‘the peasant way’; [po pros'tomu] (lit.) ‘the simple way’ (particularly in Drahičyn); [po 'naʃomu] (lit.) ‘our way’; and [po 'silskomu] (lit.) ‘the village way’ (in Luniniec). The name of the village is also commonly used to refer as a hyponym to describe both the local variety and West Polesian; e.g. [po bɔdɐ'nɪʊskomu] in Bahdanaŭka (Luniniec), [po 'tɔlkɔʊskomu] in Tolkovo (Drahičyn), [na ʒɪd'tʃanskomu] in Žydčə (in Pinsk). Other less common names which I have

³ *Litvyny* is an obscure term which literally translates as ‘Lithuanians’ in the sense of the historical Great Duchy of Lithuania, which was composed of more ethnicities than contemporary Lithuanians: primarily Jews, Belarusians and Poles.

documented include (in Tolkovo) [po do'maʃnʲemu] (lit.) 'the home way'; [po ko'lisnʲomu] (lit.) 'the old way'; [po mʲas'tetʲnomu] (lit.) 'the local way' (in Kamieniec); and [na 'naʃɪm jazɪ'kovɪ] (lit.) 'in our language'.

In my field experience I can see that speaking West Polesian carries a stigma for most, and for that reason, many struggled to “open up” and use it when I was around. In some villages more than others, parents have actively avoided speaking West Polesian to their children in order to save them from the stigma and hoping to give them a better future. However, very often these children learn West Polesian from their peers in the village and can code-switch between Russian and West Polesian. Most of the people who have left to go to the city very consciously avoid using it (many even claim they have forgotten it, although I have noticed some of them speaking West Polesian to their mothers or grandmothers). People who have moved to the larger cities hardly ever pass the language on to their children. In fact, several adults settled in urban areas shared with me the experience of being told off (or “corrected”) by their own children when they heard them speaking West Polesian.⁴ In the best cases, people who have emigrated to the cities and their children looked at West Polesian culture (and my work) with certain exoticism and humorous interest. Yet hardly any of them have reported any interest in undertaking the study of their relatives' language.

⁴ The best exception to this, so far, was the town/city of Drahičyn (circa 14,000 inhabitants), where I have observed a high degree of bilingualism/diglossia, although Russian is dominant.

1.2.2. History of Belarusian

Standard Belarusian (BLM) and West Polesian are significantly different from each other, yet their stories overlap at different points and both of them have been quite marginalised through history.⁵ In fact, until the late fifteenth century, differences between Eastern Slavonic varieties (even within the present standardised or recognised ones) were minimal. For this reason, I will briefly introduce the history of BLM.

Lomtev (1956) believes that the first written texts containing features characteristic of Belarusian dialects date from the thirteenth century, during the times of the Great Duchy of Lithuania. The presence of “(Old) Belarusian” kept slowly increasing until the sixteenth century when its use (other than for personal correspondence) became fairly widespread following the publication of Skaryna’s translations of fragments of the Bible.⁶ However, the Polish rulers of the Great Duchy of Lithuania gradually kept excluding Belarusian, culminating with a decree in 1696 that banned writing and publishing in Rus’ian (i.e. Eastern Slavonic, Belarusian/Ukrainian/Polesian) (Lomtev 1956: 5). In the nineteenth

⁵ Although BLM has gained recognition during the twentieth century.

⁶ Note that the language into which Skaryna translated the Bible still provokes intense political and even academic debates nowadays. I do not wish to enter into such debates and I lack enough evidence to defend a specific position. Yet, so far, I can point out some of the commonly agreed facts. Eastern Slavonic varieties were not as distinct from each other as they are nowadays. The text was written in a late Common Eastern Slavonic (or Old Rus’ian), and it definitely sounds obscure and arcane to any modern day (untrained) Belarusian speaker. Yet, already at this stage, some of the differences between Eastern Slavonic varieties were noticeable. So according to Lomtev (1956: 4-5) some of the particularities of Belarusian (but not West Polesian) could be observed in Skaryna’s writings, such as non-palatalised /r/ or *dzekanne* (i.e. /d / + /j/ = /dz/).

century, Bahušëvič started proclaiming the dignity of and advocating the use of Belarusian language and identity. However, it was not until somewhat later that mainly Janka Kupala, Jakub Kolas and Maksim Bahdanovič settled the standard for Modern Literary Belarusian, based on the dialects from Central Belarus (i.e. around Minsk). With the victory of the October Revolution, Belarus acquired its independence. The new government strongly encouraged the use of Standard Belarusian (based on Taraškevič's (1918) grammar)⁷ in the administration and in education by founding Belarusian schools and the Belarusian State University in 1919. This caused other varieties (particularly West Polesian) to be even more stigmatised and excluded from any educational systems or writing. Stalin's accession to power undid some of these reforms (favouring 'russification') and pushed for the creation of a new standard (known as *Narkamoŭka*) which was closer to Russian and even further from Western varieties (including West Polesian).

1.2.3. History of West Polesian

When it comes to West Polesian, I have already mentioned that the area of Western Polesie has been quite isolated for centuries due to the topography of the area. Isolation has contributed to the conservation of some cultural and linguistic features from the Common Slavonic era (such as a de-obligative future tense form), even though West Polesian also presents some innovations (e.g. pronominal numerals).

⁷ Which gave origin to the standard known as *Taraškeviča*.

With the appearance of (paved) roads connecting the villages to the cities, the population has increasingly been exposed to the languages spoken in the city (in the case of my study, Russian). In addition to this, the oldest generation of speakers born during Polish rule (1917-1938) had their schooling (up to third or fourth grade, when possible) in Polish. A tiny minority who were able to attend school during the Nazi occupation reported being schooled in Standard Ukrainian for one or two years. Nowadays children are schooled in Russian or Standard Belarusian, and the media are most frequently in Russian (although they can also listen to programmes in Ukrainian, Belarusian and even Polish).

Religion is an important component of the cultural identity which also has an effect on language preferences and use: Russian Orthodox believers attend services in Church Slavonic, although the teaching and the communication with the priest are done in Russian. Protestant services are conducted in Russian, although they are open to songs and preachers who speak Standard Ukrainian and Standard Belarusian. Yet, since lay people (who often have no higher education) do most of the preaching, the Russian they use is frequently full of Polesian words, structures and phonology. Catholics go to mass in Polish and/or Standard Belarusian, depending on the parish.

All this is leading to changes in grammar and lexicon, and even to the language not being used by the younger generations migrating to the cities. Being surrounded by other standardised languages (i.e. in the cities or at school), which are genetically very closely related, the area could be regarded as crossroads

where different morphological and syntactic systems clash. In fact, Korjakov (2002) described the situation in Western Polesie as diglossic.⁸

The UNESCO Atlas of Endangered Languages (2011 online version) listed “Polesian” as one of the endangered languages in Belarus.

“Only two East Slavonic languages, Russian and Ukrainian, are classified as non-endangered here, meaning that Belarusian, while an official language of an independent country, is regarded as vulnerable, based on the widespread use of Russian in its stead. Two regional languages are recognised in this group: one is well established, i.e. Rusyn [...]; the other is somewhat unknown, namely Polesian in the border region of Belarus, Poland and Ukraine, still subsumed under Belarusian by SIL and in the Encyclopedia (Moseley, 2007)” (Salminen 2010: 37) (see references there).

The first book in what can be considered West Polesian is Dunin-Marcinkevič's *Pinskaja šlaxta* published circa 1886. The book is a play in which Dunin-Marcinkevič replicated the speech of the people in Pinsk and surrounding villages. Nevertheless, it took another hundred years to see more literature being published in West Polesian. On the brink of the collapse of the Soviet Union, in the 1980s and 1990s a pro-Polesian political-linguistic activist group emerged led by Šyljahovič. Inspired by Duličenko's (1981) work they advocated a “West Polesian Literary Microlanguage”, and they created an alphabet for it (Šyljahovič 1990), which they used in their publications (Paljakoŭ 2013).

⁸ Interestingly enough, Korjakov is the first sociolinguist to deal with West Polesian and one of the first scholars to start distinguishing West Polesian from Belarusian dialects.

They created a magazine called “*Zbudinnje*” (‘awakening’),⁹ where they published articles on political and linguistic issues affecting mainly Polesie in Russian and Literary West Polesian. However, the group often (although not always) had autonomist claims, which were not particularly welcomed by the government (Paljakoŭ 2013). By the end of the first decade in 2000, the group in Belarus was almost extinct (and, with it, a part of the interest people had put in the Belarusian side). Nowadays most of the research and activism in Belarus is centred around *The Zaharoddze Civic and Academic Organization* (which aims to be politically neutral), which was led by Klimčuk until his death (on the 22nd of October 2018). Through the organisation Klimčuk published a few translations in West Polesian, the most noteworthy being his translation of the New Testament (first portions published in (2010), whole NT forthcoming).

1.3. Previous research

West Polesian grammar has been very little studied. This is partly due to the tendency of the national academies of the respective countries (except for Poland) to include it as a variety of their standard language (e.g. Avanesaŭ 1964, Bevzenko 1980, Hulickaja et al. 1992).

The pioneers in Western Polesie started their research a century ago. Among them, the three figures that deserve a mention are Obrębski, Vjarenič and Seržputovski

⁹ The publications from 1989 to 1990 are available at: <https://vytoki.net/?docs=00004971>

(1911).¹⁰ The work of all these is far closer to ethnography and the study of folklore than to the language system itself. Nevertheless, they published the first transcribed texts, and some information about geographical terms (often related to hydrography), farming and the household.

In my MA thesis (Roncero 2015) I provided a quite extensive account of the publications on West Polesian, which are not many. Now, three years after that, I cannot say I have identified any other meaningful publications on West Polesian, particularly regarding morphology and syntax.

As I already pointed out (Roncero 2015), the most purely linguistic publications deal only with the lexicon (Arkušyn 2016, Malažai & Klimčuk 1989, Tolstoj 1968) and phraseology (e.g. Kascjučyk 2005). The only meaningful exception has been Klimčuk's (1983) description of the phonology of four different West Polesian varieties. His classification of West Polesian varieties is still a reference work for people working in the area.¹¹ In fact, Klimčuk can be considered the father of contemporary linguistic studies on West Polesian grammar.

As I mentioned in my previous work (Roncero 2015), on the Polish side, there is apparently an increasing level of activism from younger generations to use it.¹² This may have had an effect on some Polish scholars engaging on the study of Podlasie

¹⁰ Actually, Seržputovski's work took place in Eastern Polesie, but still within the territory of the contemporary Belarusian Polesie. Part of Vjarenič's fieldnotes and work were published in (2009).

¹¹ Even though, as I mention in (§2.1.), the isoglosses for morphology and syntax are very different from the ones used by Klimčuk (1983), which are fundamentally based on phonology.

¹² See for example www.svoja.org

and Belarusian Polesie, such as Barszczewska (2016), Engelking (2001, 2006), Golachowska (2006) and to a lesser degree Jankowiak (Jankowiak & Grajewski 2011). Yet their work is primarily related to the lexicon, semantics and ethnography.

On the Belarusian side, there have been different publications on dialectology which covered some morphology and syntax but very superficially. Most often they have been written as a part of a bigger and more general linguistic survey (Avanesaŭ et al. 1963, Bosak & Bosak 2005, Levancèvič 1994, 2013), and they lack any analysis of the forms. Moreover, in all of the cases so far the work has been undertaken with a purely quantitative approach. For example, The Dialect Atlas of Belarusian (henceforth DABM) (Avanesaŭ et al. 1963) covers more than 200 parameters from very different parts of the grammar and lexicon of more than a thousand villages in Belarus. Furthermore, in the case of DABM the research was undertaken between 1949 and 1958, which implies three significant factors possibly affecting the results. First, the likely passing away of most interviewees and, therefore, the generation using the old system. Second, the extended influence of Russian-speaking rulers in the region (which had just been ruled by Germans and Poles). And third, the “civilisation” (as they refer to it popularly) of most of the villages involved in the study. The building of roads between 1980 and 1990, has led to greater contact with, and thus pressure from, Russian (spoken in towns and cities).

Ironically, right after returning from my first fieldtrip I came across an article by Del Gaudio (2014), in which he describes in brief outline the particularities of West Polesian. He also recognises the need for deeper linguistic research, but shows little hope of any chances to do fieldwork:

“In the current state and due to the contingent difficulty of undertaking fieldwork in the Belarusian area, it will be necessary to start working on the basis of the already existing cartographic [sic] material; by initially conducting an empirical study restricted to the Ukrainian transitional dialect area” (Del Gaudio 2014: 286) [My translation].

In addition to this, typological approaches are almost non-existent in Belarusian dialectology, or even for Standard Belarusian. One of the few exceptions is Klimčuk’s very brief comparison of West Polesian phonology with the Southern Slavonic subfamily (1973), and his concise historico-typological work (2006), where he relates the particularities of Torokanian vocalism (one of the four main varieties of West Polesian) with phenomena in other Indo-European languages.

Narrowing down the focus, more than two thirds of the core content of this thesis deal with numerals and phrases containing numerals, i.e. numeral phrases (NumPs), in West Polesian. I document their behaviour, as they are peculiar within the Slavonic family and even cross-linguistically. As I point out in Chapter 5, there is very little literature, it is hard to access and it contains many inaccuracies. Having said this, it is interesting to note that Kim (2009), who probably had little or no knowledge of West Polesian, had already pointed out a prominent data gap of this parameter in Slavonic varieties:¹³

¹³ Many authors have written about numeral phrases in Slavonic, particularly about Russian and Polish (see Chapter 4 and 5). Nevertheless, most of these works have been analyses, and only few of them provide “raw” data (i.e. from fieldwork). Furthermore, certain Slavonic varieties are particularly underrepresented, such as Kashubian, Rusyn and West Polesian.

“In addition to these syntactic issues associated with numeral phrases, i.e., DPs/NPs, more field researches on numeral phrases are needed to acquire authentic data from native speakers, who belong to a wide range of different groups of age, gender, occupation, and education. Furthermore, aside from Russian, BCS, and Polish, other Slavic languages are worthwhile to be surveyed to analyze the differences and similarities of them with each other. Especially Upper Sorbian, Lower Sorbian, and Slovene, which still have dual number, are important as sources to gather the information of historical change in the Common Slavic numeral system” (Kim 2009: 189-190).

Hence, the research gap is undeniable. Based on data from substantial and predominantly qualitative fieldwork, this thesis aims to contribute to the areas of inflectional morphology and syntax from a typological perspective.¹⁴ Moreover, I intend to do this using primarily the framework of Canonical Typology (Bond 2018, Brown & Chumakina 2012, Corbett 2005), which I explain in further detail in Chapters 5 and 6.

1.4. Writing conventions

I am aware that multiple conventions can be proposed for transcribing data from West Polesian, based on different and probably more detailed analyses of the phonology. However, as I have stated before, the focus of my PhD was not phonology, so the

¹⁴ All the work in morphology concerns exclusively inflectional morphology. I take Matthews’ (1991) definition as a basis for this work. “We may define **inflectional morphology** as the branch of morphology that deals with paradigms. It is therefore concerned with two things: on the one hand, with the semantic oppositions among categories; on the other, with the formal means, including inflections, that distinguish them” (Matthews 1991: 38).

conventions presented here must be understood more as an impressionistic transcription and a working orthography.

There have been several attempts to create an orthography for West Polesian. There are three main orthographic conventions. The first one is Šyljahovič's (1990), which was used mainly for his journal "Zbudinnje" (a modification of the Belarusian and Ukrainian Cyrillic with some innovations such as the grapheme <j>). The second one is Klimčuk's (based on Belarusian Cyrillic), which he used in his translations (most notably the New Testament portions (2010)). The third one is Maksymiuk's (2007) orthography for Podlasian (a Latin script based on Polish, but with a few additional diacritics), used by Podlasian activists blogging and creating memes.

For the sake of clarity, and in order to make this research as accessible to as many linguists as possible, I have decided to use the International Phonetic Alphabet to transcribe West Polesian. There is a long tradition of Latinisation for other Slavonic languages, most notably Russian, and so there are already some conventions established (admitting that there are different standards for this as well). Nevertheless, the most common conventions (and the one I will use when transliterating Belarusian, Bulgarian, Macedonian, Russian and Ukrainian in this work) do not correspond to the standards of the IPA, which does not facilitate the work of the general linguist when going through a large sample of grammars.

In any case, for practical reasons the phonemes /t̪/; /d̪/ and /l̪/ will be represented as /t/; /d/ and /l/ respectively.¹⁵

Even though I am still uncertain whether it comprises a “full phoneme” in West Polesian’s inventory, when I have identified some speakers using a glottal stop /ʔ/ I have represented it. The context where the glottal stop has appeared is when /b/ is between homorganic vowels, most commonly (if not exclusively) in the word /treba/ ‘it is necessary’, which in Torokanian varieties is pronounced as [ˈtraʔa].

In Standard Belarusian there is a phonological rule by which /v/ becomes /ǔ/ when preceded by a vowel (but not after a break). In West Polesian, the corresponding alternation is between /v/ and /ʊ/. Yet it is not always clear that all speakers apply the same phonological rule, so I have opted for transcribing what I heard however inconsistent it may seem.

The phonemic function of palatalisation in West Polesian is less clear than in other Slavonic languages. In any case, I have decided to render palatalised consonants as the combination C + j (e.g. [losʲ] > /losj/ ‘elk’). See more on palatalisation in (§3.1.3).

In spite of the IPA standards, I use capital letters/graphemes when a proper noun is employed for the sake of clarity; e.g. ¹Saʃa.

¹⁵ The velarised realisation of /l/ changes considerably from one speaker to another and it is not always easy to hear it (i.e. often it seems to be realised as non-velarised). In any case, I have tried to distinguish it from the “palatalised form”. Thus, as a convention when /l/ is followed by /a/, /e/, /o/, /u/ or /ɪ/ will represent [l̪] (even though this velarisation was not clear from the recording); and when /l/ is followed by /i/, /j/ or any diphthong headed by /j/, it will represent [l] or [lʲ].

I mark the stress <'> on the nouns when it is relevant for the inflection. In the majority of instances, the present tense and the imperfective aspect are the unmarked forms in the verbal bases. For this reason, I only indicate them in the glosses when they are not the default form (e.g. when they are marked by an infix). In the same vein, given that cardinal numerals are the most common (and unmarked), I do not specify the type of numeral in the gloss if it is a cardinal numeral.

I transliterate standardised Slavonic languages with a Cyrillic alphabet according to the Slavonic and East European Journal (SEEJ) conventions in Comrie & Corbett (1993: xvii-xviii). Nevertheless, I use the official transliteration for place names in Belarus (State Committee of the Property of the Republic of Belarus 2007), “On alteration and amendments to the Regulation of geographic names of the Republic of Belarus transliteration with the letters of Roman Alphabet”, which was later proposed to and adopted by the United Nations (2007).

When transliterating Common Eastern Slavonic, Old Belarusian, Old Church Slavonic, Old Russian and Old Ukrainian texts I follow some special transliteration rules, which are commonly used in the discipline (see Cubberley 1993: 57-58), and I may use a specific font:

<г> = <g> Russian (including Old Russian), Bulgarian, Macedonian, Bosnian-Serbian, OCS; <h> Belarusian and Ukrainian (including Old Belarusian and Old Ukrainian).

<у/ѳ> = <u>

<ѡ> = <o>

<ѣ> = <ě>

<ѣ> =

<ѣ> =

1.4.1. List of abbreviations and glosses used in this work

I present glosses and abbreviations according to the Leipzig Glossing Rules (following the version updated in 31/05/2015).¹⁶ In addition to the most frequent ones, I have had to create abbreviations for very frequent terms, where possible adjusting to the most frequent form in English academic literature:

1: first person	COLL: collective numeral
2: second person	COLS: collective numeral substantives
3: third person	CSR: Contemporary Standard Russian
ABS: absolutive	CT: Canonical Typology
ACC: accusative	CZ: Czech
ADNM: adnumerative	DABM: Dialect Atlas of Belarusian
ART: article	DAT: dative
AUX: auxiliary	DET: determiner
BCMS: Bosnian-Croatian-Montenegrin-Serbian	DIS: distributive numeral
BG: Bulgarian	ERG: ergative
BLM: Standard (Literary) Belarusian	EMP: emphatic
CARD: cardinal numeral	F: feminine
CES: Common Eastern Slavonic	FUT: future
COMP: complementiser	GEN: genitive

¹⁶ <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>

GRADNM: greater adnumerative	PL: plural
HR: Croatian	PRON: pronominal numeral
HSBM: Historical Dictionary of Belarusian	POL: Polish
IMP: imperative	POSS: possessive
INS: instrumental	PRF: perfective
IPFV: imperfective	PRS: present
LCS: Late Common Slavonic	PST: past
LOC: locative	Q: question particle/marker
M: masculine	QADV: quasi-adverbial numeral
MKD: Macedonian	REFL: reflexive
N: neuter	REL: relative
NEG: negation	RU: Russian
NOM: nominative	SG: singular
NONSG: non-singular	SK: Slovak
NP: noun phrase	STAT: status numeral
NumP: numeral phrase	SVO: Slovene
OBL: oblique	ULM: Standard (Literary) Ukrainian
OCS: Old Church Slavonic	US: Upper Sorbian
PART: particle	VOC: vocative
	WP: West Polesian

Chapter 2

Methodology

As I have already explained in the previous chapter (§1.3.), West Polesian morphology has never been systematically studied, nor are typological approaches common in Belarusian linguistics. For that reason, I had to collect information through linguistic fieldwork. The data obtained results from spending almost eight months in Belarus, divided over three expeditions. I interviewed and gathered oral texts from fifty-three native speakers; although I did most of the work (especially grammar elicitation), I did have the assistance of two fluent language assistants: B6 and Tor1. Overheard conversations also played an important role in acquiring the language and confirming some of the speakers' grammaticality judgments. Once the data was gathered, I thoroughly played through all the recordings at least twice, searching for specific phenomena, which I then carefully transcribed. In light of this data, I also elaborated and expanded the questions within the questionnaires for my next expeditions, in order to study some of the parameters covered in this research in greater detail.

In this chapter, firstly, (§2.1.) I explain how I chose the villages and the language assistants for this project. Secondly, (§2.2.) I provide details about the procedure of data collection and management. Thirdly, (§2.3.) I present some of the limitations encountered in this project. Lastly, (§2.4.) I give an overview of the ethical issues surrounding this research and conclude, (§2.5.) with a summary of this chapter.

2.1. Criteria for choosing the villages and the speakers

As part of my MA thesis (Roncero 2015), I had produced a list of all the villages where at least some sort of fieldwork on West Polesian had been undertaken in the past and when, so I could propose ‘virgin’ villages for a future expedition. Once in the field, I experienced difficulties finding contacts in those villages I had planned to investigate. Meanwhile, I started encountering other more interesting places to work, as the phenomena I wanted to study revealed themselves. As a result, I devoted a considerable amount of time in the field identifying new potential villages, making contacts with people who could host me in situ, and getting the permits to visit them (see more details in (§2.3.)).

The criteria I used for selecting the villages were the following:

- Villages where there is still a considerable number of speakers of the local variety.¹⁷
- Villages that were representative of at least one of the four main variants of West Polesian, based on Klimčuk’s (1983) and Levancèvič’s (2013) maps, as well as their personal comments.
- Villages where there had not been any previous fieldwork done (at least substantially), as a way to bring more diversity to the existing materials on West Polesian.

¹⁷ To my surprise, many villages had been largely populated by people from other regions of Belarus and Ukraine in the last years.

- Villages where I could find a trusted host (healthy enough to take care of themselves); and who could help introduce me to the first contacts. These people were usually local social workers or the *starosta* ('manager') of the village.
- When possible, villages which are (or have been) geographically isolated and/or transition points (with Ukraine).

Regarding the existing linguistic maps of the area, I tried to find out what the criteria were for such classification (i.e. just phonology, or also lexicon, morphology, etc.), before paying more attention to their divisions. My initial intuition was that they had been created mostly based on phonology (in part, because to my knowledge nobody had dealt systematically with West Polesian morphology or syntax). This intuition was confirmed later on, as the phenomena in morphology and syntax that I had been studying had proved to follow other geographic divisions; roughly, western varieties (where, for example, I attested a rich inventory of future tense constructions and suppletive stems for year, or postnominal possessors) vs eastern varieties (where I documented most of the pronominal numerals and adnumerative forms, but hardly any suppletive stems for 'year').¹⁸ Thus, by the end of the first expedition, I decided not to pay too much attention to them anymore, as they did not serve my purposes, other than just knowing which varieties were not already part of the West Polesian continuum (such as Eastern Polesian or North Brestian).

¹⁸ East (Žydčā, Bahdanaŭka and, less certainly, Vostraŭ-Pare) vs West (Tatarja, Tolkovo, Khabovičy, and surrounding villages). See more details in (Appendix II. Map of villages covered in this research).

2.2. Data collection

The initial idea was to base the research on qualitative methods. In this section, I explain my methods for conducting the elicitation sessions (§2.2.1.); how I managed the data and present it in this work (§2.2.2.); and I give details of the equipment used for recordings (§2.2.3.).

2.2.1. Data-elicitation sessions

My initial approach was to observe the language in free texts, as I had not had any (significant) previous contact with it. Once I had singled out some of the parameters I considered interesting for study purposes, I started narrowing the focus of my research and designed questionnaires that were more specific (e.g. within the topic of adnumerative forms, their interactions with adjectives and ANIMACY).

I researched more phenomena than the ones covered in my thesis. Some examples of the topics I elicited, but which did not translate into any chapter in this thesis, include: the ongoing DATIVE-LOCATIVE syncretisation; heteroclisis in nominal paradigms; the VOCATIVE case; and PERSON marking in the overt copula.

I was familiar with different forms of elicitation and their contexts, but I tried favouring text-based approaches (e.g. Thieberger 2012) for linguistic research. In order to mitigate speakers' discomfort with being recorded and ensuring a large degree of naturalness, I used prompts (e.g. asking a question about familiar topics), visual stimuli (i.e. wordless books inspired by the ideas in Sake & Everett (2012)) and problem-solving/creative tasks.

The choice of books that would suit the cultural context proved to be appropriate. Nevertheless, developing tasks to help the elicitation did not always provide the expected results. In particular, some of the tasks demanding a bit of abstraction or creativity failed to provide any meaningful results with some very old people even though I had explained the task and given examples of it. For example, in order to study the lexicalisation of a possible DUAL, I drew images of people with uneven numbers of ‘natural pairs’ (e.g. three arms and five eyes) and asked them to tell me what were these people’s special abilities (e.g. ‘he has five eyes, so he can see things in the distance’). Moreover, most of the parameters I was studying appeared very rarely (especially the FUTURE TENSE). That created many blanks in the paradigms, besides all the extra noise generated by this method. As a result, not only was I unable to study the syntactic behaviour of very specific (and rather marginal) constructions, I was also not able to properly gather the entire paradigm of most lexemes solely based on the free text corpus. That led me to a reflection on a restructuring of the methods and approaches, so I ended up adding questionnaires with direct-elicitation tasks to the work sessions with the language assistants.

The type of tasks and questions I had initially foreseen were suitable for a more qualitative approach. However, based on different speakers’ critiques of their neighbours’ recordings I experienced confusion about which forms were genuinely West Polesian or a calque. As a result, I started asking different types of questions, which got more reliable responses and lent themselves to being counted. Thus, at some points, the research turned more quantitative than qualitative (especially in Bahdanaŭka, where I interviewed up to twenty people); and so, the corpus of this

research is based on interviews of fifty-three native speakers.¹⁹ As a by-product, involving many speakers helped to reveal the high amount of variation in certain parameters, which contributed enormously to Chapters 5 and 6. Yet, I did obtain a substantial amount of negative data which contradicted my initial suspicions and some of the language consultants' recurrent comments (such as that there is no division between 'lower higher numerals' and 'greater higher numerals'; or that there is no discontinuity in morphosyntactic government and/or agreement within the group of lower numerals).

The workflow of the elicitation sessions with the language assistants largely followed this structure. I dedicated the first minutes to eliciting one or two free texts, sometimes I would suggest to them a specific question about their life (e.g. a mischief they did as kids). However, often, those who had been working with me for a while proposed the topics themselves. After the free texts, we would move on to prompts. For example, when I was working on the FUTURE TENSE, I would give them a wordless book and ask them to retell me the story as something they were going to do on the next day. Often this generated more noise than expected, so I selected a few images and asked new and more specific questions (e.g. "how many cows do you see?", when eliciting adnumerative forms). Finally, I moved to more direct elicitation, following a questionnaire of very specific questions I had elaborated before the session. The direct

¹⁹ This is also partly caused by the fact that not all the speakers I wanted to interview were always available. Most of them were only interviewed once, as I had not met them before, and I did not know how suitable they were for such tasks. Men in particular tried to use Russian or a very 'russified' speech with me, which in most cases made me disregard the possibility of working with them in the future.

elicitation involved some grammaticality judgements as well as paraphrasing sentences.

The language of contact for the sessions kept changing. I designed my first questionnaires in standard Belarusian. However, some speakers would immediately complain that they did not speak or even understand Belarusian. I had to use Russian with them at some points, although I avoided it as much as possible because it provided unnatural and unreliable responses. Often the speakers concentrated so much in providing a 'word-to-word translation' rather than a paraphrase that they even ended up denying using genuine West Polesian constructions which I had heard them using previously. Thus, once I became more fluent I started conducting the sessions in West Polesian.

I also systematically kept track of the development of my interviews using a template I created (<http://doi.org/10.5281/zenodo.1118964>). I used this to record any factors that may have altered the interview (e.g. mistrust; health issues, etc.). In addition, I used the template to correct and improve my performance as a fieldworker over time.

Besides the materials elicited during the sessions, overheard conversations were a highly valuable source of information. I spent a significant amount of my time in the field observing the interactions of the speakers outside the elicitation sessions in order to check the accuracy of some of their judgments and responses provided during the sessions, and of course acquiring the language. When I was allowed to do so, I undertook participatory observation, which most often consisted of farming. Such a task ended up being crucial for understanding their worldview (i.e. their conversation

topics and the texts they shared during the interviews), and which also contributed new ideas for the questionnaires.²⁰

Another important part of my fieldwork was finding the best language assistants. In order to do that, I dedicated a decent amount of time trying to become visible (mainly by going to the local *banjas*;²¹ going to the local shops as often as possible; and attending church services). Becoming visible often translated into more people being comfortable to work with me, and in some good cases, my gaining their trust.

2.2.2. Classification of the data

I have been able to record a considerable amount of data. Even though I had listened to most of the recordings in the field (and made notes), I spent several months after that listening to them again. Whenever I heard any of the constructions I was looking for, I transcribed the utterance (with a bit of context), followed by the name of the file, and the minute and second in the recording. Nevertheless, the research questions changed several times during the first expeditions and the months after, as some phenomena revealed themselves to be more interesting than others.

Given that the ‘naturalness’ of the utterances can have an effect on the reliability of the data, for the sake of transparency I have decided to tag the examples according to the following code (inspired by Sakel & Everett (2012):

²⁰ In general, when I started eliciting sentences with a certain structure, using a very specific topic about their way of life (about which they are used to talk in West Polesian), the results were generally more accurate and the tasks were better understood.

²¹ Traditional Eastern Slavonic steam room.

- **EL:** Directly elicited or suggested form (i.e. grammaticality judgment).
- **OV:** Overheard, outside the context of elicitation.
- **PR:** Prompted, elicited indirectly (i.e. suggesting a context).
- **VOL:** Volunteered by the speaker.

Other than those stated, if there is the number of the recording followed by the time, the example has been directly retrieved from the corpus of transcribed free texts.

2.2.3. Technical equipment

The vast majority of audio files were recorded using Marantz PMD661MK and PMD671 solid state recorders in *.wav format (48Hz, 16bits). All the video materials were recorded using a Canon XA10 (HD) camera, with an *.mts output format.

The recording devices were complemented with external microphones: an AudioTechnica (AT897) unidirectional condenser (or shotgun) for the majority of video and audio recordings and a Sony (ECM-MS957) stereo microphone for some video recordings.

2.3. Limitations

As was to be expected, I faced different limitations undertaking my research which had an impact (some greater than others) on the project. On the one hand, I had to deal with practical issues associated with travelling to the field and being able to work there (§2.3.1.), as well as funding the expeditions (§2.3.2.). On the other hand, I had to deal with a complex sociolinguistic context of diglossia (§2.3.3.), as well as purely sociological factors conditioning the sample of participants, gender being the most significant (§2.3.4.).

2.3.1. Time, geographic and legal limitations

Due to time and physical limitations I was only able to cover some villages (representing different varieties of West Polesian) in the region of Brest, in Belarus (see (*Appendix II.*) for more details). Many of those villages were difficult to access as they had a very restricted transport service, if there was any, or the roads were not good (either flooded or with dangerous animals) which was an ongoing difficulty during the three expeditions. Moreover, I had to be very careful with the levels of radiation in some of the villages, and consequently avoid some of them. For example, I wanted to work in Pare (Pinsk district), but most of the population had been evacuated after the accident of Chernobyl, due to the high levels of radiation. As a result, I was staying 7km away from the village (in a safer village), although with very restricted means of transport to reach it.

Consular and legal issues have also been a very big burden. Already before starting my PhD I had been making arrangements for my fieldwork expedition to Belarus,²² and I also dedicated a significant part of the first months of my PhD to that enterprise. However, the bureaucratic procedures kept being delayed for various reasons, so I had to wait almost seven months to obtain a visa invitation letter for my first expedition. Once in Belarus, I thought I had solved most of the problems, but actually, I wasted more time than expected with additional procedures (which included applying for special permits to travel to villages that are less than 30km from a border

²² I actually did travel seven months before on a tourist visa in order to find out information about research visas; find a hosting/inviting institution; explaining my project to them; and starting the visa application process.

zone). That interrupted my research considerably, as I was obliged to return to the city from very distant villages in order to deal with their documents on several occasions. Once in the villages, I devoted a considerable amount of time gaining local people's trust (starting by 'becoming visible')²³ and identifying potential speakers. Once I had gone through hardship on the first expedition, the second and third expeditions ended up being more time-effective.

2.3.2. Funding

The studentship I obtained for this PhD project did not cover fieldwork expenses, thus I had to raise my own funds. I approached many charities and trusts, but only three of them gave a positive reply. Thus, The Gilchrist Educational Trust, The Philological Society (GB) and the Belarusian Charity (GB) supported my first and longest fieldtrip to Belarus, for which I am very thankful (see (*Acknowledgements*)).

The second and third expeditions were shorter in time and only covered one parameter and village at a time. Moreover, since I had to cover the cost of the trip on my own (since the other funding bodies could only provide support once I had entered into the actual PhD), I had to limit the time of my stays.

2.3.3. Standard language(s) bias

This has been one of the giants I have had to be constantly battling against. West Polesian is an Eastern Slavonic variety which is closely related to all the neighbouring

²³ I dealt informally, although in more detail, with this topic on this entry from my blog <https://westpolesian.wordpress.com/2016/10/25/preparing-for-fieldwork-i-visibility/>

languages or varieties WP speakers are exposed to (to a higher or lesser degree). These varieties are quite strongly standardised (some more than others), whereas West Polesian is primarily used for oral communication with family and people from the village.²⁴ As a result, speakers often lacked awareness of what was ‘correct’ or ‘belonged to the system’ and what did not. In addition, there is not a neat perception of national identity or language, as we would expect from other Western highly standardised, educated, industrialised languages/nations (see more on this in (§1.2.)). As a result, very frequently it was hard to determine whether a certain form could be properly classified as West Polesian, a borrowing, or code-switching (depending on the speaker’s knowledge of other varieties/languages). A common factor affecting the reliability of many of the elicited materials has been people trying to use Russian with me (for being an outsider).²⁵ I only found out that a participant was using ‘pseudo-Russian’ or speaking very differently from how they usually did, after other people from the village had heard the recordings and commented on them. My main criterion in these situations was to trust the comments of other speakers so as to assess the legitimacy of the collected forms. On top of this, whenever I encountered ‘new’ or ‘exotic’ constructions, I checked the grammars of all the surrounding languages to see whether there could be any relation between them (e.g. calque).

²⁴ In fact, I noticed, that when a West Polesian speaker encounters a speaker of a different village, that is not in the area (even if their varieties are closely related) they tend to communicate in Russian (even though, they may not be native speakers).

²⁵ Except for a couple of speakers, especially one who used Standard Ukrainian words deliberately to avoid Russian loanwords, as a sign of distinctiveness (and pro-Polesian activism). Jankowiak also reported a tendency to value Russian to the detriment of the local variety in Latgale: “Pieces of conversation with my respondents show that, for locals, Russian definitely has greater prestige than the Belarusian dialect [...]”(Jankowiak 2014: 353).

Many linguists working in Belarus as well as some (non-linguist) native speakers shared with me similar concerns about everyone's variety being increasingly affected by Russian. This overlap of different closely related language systems produced a fair amount of variation, even within the same village or the same speaker. Jankowiak (2014) reports a very similar situation after having worked on Belarusian dialects in Latgale:

“Depending on the level of fluency in the other language/dialect, there is a different level of language interferences in the speech of each individual. This results in a significant idiolectal diversity within a dialect on an idiolectal level. Idiolectal features, as noted by Smułkowa, are typical of mixed dialects.

The material collected indicates great diversity in the informants' speech, not only on dialectal, but also on the individual level. Here we can talk of a high level of idiolectal differentiation. A few minutes of conversation are enough to notice that the interlocutor either uses two or more words (Belarusian and Russian) to describe one lexeme, or one word is realized once in Belarusian pronunciation and another time with Russian phonetical [sic] features[...]" (Jankowiak 2014: 352).

In any case, I will deal with variation in further detail further on this work, particularly in (Chapters 5, 6).

2.3.4. Gender imbalance

From the whole corpus, only fifteen men (27.77%) took part at least in some way, in contrast with thirty-nine women (72.22%). Besides, men's interventions were considerably shorter than women's (often participating as 'people in the back').

None of the main language assistants I worked with was a man, although B9, B6's son was in all of my interviews (except for one), and generally provided good explanations or corrections to his mother's 'mistakes'. B5, B6's husband was also present in all of our interviews, but he only intervened a few times (except for my initial interview with him).

There are several reasons why this is the case:

- One of the main reasons is that life expectancy for men is considerably shorter than for women in the area; in fact, a large number of the women I interviewed were widows. This was related to several factors: high levels of alcohol consumption in the area (especially among men); many men died or were severely injured during WWII; and a high rate of thyroid and heart related diseases existed in the area.
- In rural areas most men under fifty-five worked in jobs outside of their village. They would often leave the village for seasonal work (mostly in the building industry) in bigger cities; typically Moscow, Minsk or (less frequently) Warsaw. Thus, during certain periods of the year (from early spring until approximately mid-October) it was very hard to find young men (aged below fifty-five) in the villages that were free for interview.
- As a result of working in bigger Russian-speaking cities, or often leaving the village to do their military service (strictly in Russian), men's speech is very often more affected by Russian than women's, to the point where some of the men would use predominantly or even exclusively Russian in all their

interactions, as a sign of status. Consequently, whenever I asked people in the village to take me to good speakers, they would far less frequently recommend me to visit men than women (besides the other two causes described above).

2.4. Ethical issues

This project had its centre in the community, and thus I have tried to involve, assist and honour the speakers and the language as much as possible with the highest ethical and legal standards possible. This project had to undergo assessment by the Surrey University Ethics Committee (UEC) in order to receive ethical approval. The UEC's concerns turned around three areas: the recruitment of participants; ensuring the privacy of the people recorded; and the compensations. After a relatively long process of application, where I had to resubmit my application twice (due to interdisciplinary misunderstandings), the project received a favourable resolution in December 2015.

2.4.1. Obtaining informed consent

Since the community to be studied was larger than a few hundred, and part of it is scattered, it was impossible to get a collective community agreement to study or disseminate their language, thus informed consent was taken on an individual level. In order to obtain informed consent, I elaborated a form in English, which I then asked to be translated by a local into (simple) Belarusian. The written form was submitted to the University's Ethics Committee. Nevertheless, the written forms happened to be problematic; not everyone could understand Belarusian or was even literate (especially amongst older speakers). Moreover, signing documents turned

some speakers off from participating as they were scared it would be used as charges against them, because that reminded them of past experiences (during the hardest year of Stalin's repression). Thus, by the end of the first fieldtrip I started recording oral consent, whilst also giving them a hard copy of the Participant Information Sheet. Thus, I made participants aware that at any time they may withdraw from the project and, if they wished, they could have their recorded data deleted.²⁶

2.4.2. Recognition of the contribution of the participants

I have tried to give as much credit as possible to the speakers, not only by including their names in the recordings and academic publications (including oral presentations); but also through social media. However, not all of them have allowed their identity to be disclosed or in making their information publicly available. See the full list of participants in (*Appendix I*).

2.4.3. Dissemination of results and archiving

During my three field trips I collected about twenty hours of recordings. I am archiving the recordings in an academic archive, The Language Archive (TLA),²⁷ which has generously offered to host my data free of charge.²⁸ Furthermore, I have tried to disseminate as many recordings as possible on an *ad hoc* created YouTube

²⁶ Only on a couple of occasions I was requested to delete all the recordings from my recording devices.

²⁷ <https://tla.mpi.nl/>

²⁸ In order to ensure the sustainability of such recordings in the archive, I am working on converting them into 'open formats' (Good 2011).

channel (a well-known platform not only in the area, but also globally).²⁹ All this has been carried out respecting common principles of fieldwork ethics and data ownership.³⁰

Excluding those that merely contain tasks from the grammar elicitation sessions, the main reason why the number of publicly accessible recordings is still so limited is related to the ethical implications of sharing a vast proportion of them (see (§2.4.4.)). Yet, whenever it has been possible, I have shared the files with the speakers (or their relatives) who have requested so, on USB devices.

I also created a blog where I have been sharing information about the local culture, a short summary of my discoveries and my experiences in the field.³¹ In addition, I have complemented the work on the blog with updates about the project on Twitter.³² All these initiatives contributed to gaining trust and support from the local authorities as well as the attention of the press, which was useful to get more invitations to work in new villages and promote the platforms I had created for the project.

2.4.4. Additional ethical measures to ensure human dignity

No recordings that may potentially compromise a person's dignity or which contain controversial or embarrassing comments (about themselves or other people) have and will ever be publicly displayed, even if the participant may have agreed to do so.

²⁹ <https://www.youtube.com/channel/UCrxP9ybNZCmwBIATLycm8Cg/>

³⁰ Woodbury (2003: 47) defines 'the Principle of Data Ownership' in these terms: "data should not [be] disseminated to those its owners or producers do not want to have or use of it".

³¹ <https://wordpress.westpolesian.com/>

³² With the hashtag *#westpolesian*

Recordings containing inappropriate comments were very frequent (from gossip about neighbours, to stories about domestic violence). When there has been the smallest doubt, I have “erred on the side of caution” by not making the material openly available (see Macri and Sarmento’s (2010) principle).

In order not to unnecessarily restrict the access to most materials, I have followed a code for tagging the recordings, varying according to the severity of the content (with a short explanation of the cause) in the metadata: from small issues which can be solved with a bit of editing (e.g. scratching) to very unsuitable recordings for public access.

2.4.5. Remuneration for language consultants

Each participant was offered monetary compensation significantly higher than the equivalent hourly rate pay of a local teacher (4 USD).³³ Until the introduction of the New Belarusian Ruble (BYN) in July 2016, the Belarusian Ruble (BYR) had recently undergone a strong devaluation,³⁴ so foreign currency, most commonly USD, was (and is still) used as a reference for payments. Hence the economic compensation was paid in BYR according to the daily exchange rates for USD. However, in spite of my insisting, many participants refused to accept the compensation and felt uncomfortable. As a solution, I started making small gifts to them in the form of souvenirs and food.

³³ Which roughly in 2016 (before Brexit vote) corresponded to £3. According to friends working on the educational sector, the pay rate in 2015 was 2 USD/hour for new staff.

³⁴ From €1 = BYR 4,000 when I first visited the country in 2009; to €1 = BYR 22,500 when I went for my first expedition in 2016.

2.5. Summary

In sum, I have aimed for the most naturalistic speech as possible in the friendliest environment possible. Nevertheless, in order to complete certain blanks regarding infrequent phenomena it was necessary to arrange some direct or semi-direct elicitation. I have explained that I had an initial plan of the varieties I wanted to cover, based on older dialectological maps, but which did not happen to be relevant for syntactic and morphological phenomena. Beyond the most ‘practical’ limitations affecting the expedition, my main battlefield was the strong interferences from the neighbouring Slavonic (standardised) varieties and deciding which content or forms were ‘genuinely West Polesian’. Finally, I have shown that I implemented several research ethics principles that go beyond the requirements of the UEC, aiming to make a positive impact in the community.

Chapter 3

Typological overview of West Polesian

West Polesian is part of the Eastern Slavonic subfamily, and for this reason, most of the comparisons in this work are within this family.³⁵ Nevertheless, West Polesian also has phenomena that are shared with the neighbouring Slavonic varieties, including Polish; such as the VOCATIVE MASCULINE *-u* (e.g. (B7) *dzjatku!* ‘uncle’) or *-e* (*Jakob* > *Jakobe!*) (B9) (Roncero 2016); and the “Torokanian-type” (Klimčuk 1983) of vocalism [Common Slavonic] *ѣ (ě) > a (e.g. (P2) *pěsokъ* > *pasok* ‘sand’).

In general, West Polesian morphology and syntax are not very different from the rest of the Eastern Slavonic family. Through this dissertation I will be narrowing the focus to those phenomena that make West Polesian different from the rest and which are of typological interest from a wider cross-linguistic perspective. In this work I show that West Polesian also has had a unique development of certain morphological and syntactic phenomena within the Slavonic family (e.g. pronominal numerals (§4.4.)). Before going into specific parameters, I will provide a brief overview of the grammar in the light of Eastern Slavonic, whilst highlighting some of the particularities of West Polesian. In (§3.1.) I introduce the phonemic inventory of West Polesian, give a brief overview of the main phonological and phonotactic rules and compare them to other Eastern Slavonic varieties. In (§3.2.) I present the morphology of nouns (where I present some of the problematic CASE values) and verbs in West Polesian. In (§3.3.) I

³⁵ Particularly Contemporary Standard Russian (CSR) and Standard Belarusian (BLM), as they are the languages I am most fluent in.

give a short overview of the syntax of West Polesian, focusing on word order. In (§3.4.) I introduce the phenomenon of ANIMACY and point out some of the problems that the traditional analysis for Eastern Slavonic has in describing the reality of West Polesian. Finally, in (§3.5.) I give a short summary and draw conclusions in the light of the data.

3.1. Phonology

I have already explained in (§1.4.) that I adapt IPA representations (for orthographical convenience) for presenting my data, and I have already introduced some of the conventions for transcribing certain phonemes. In this section I introduce the phonemic inventory (§3.1.1.); my analysis of palatalisation (§3.1.2.); and the function of the stress (§3.1.3.). Finally, I give some glimpses of the general phonotactic rules in comparison to other Eastern Slavonic varieties (§3.1.4.).

Klimčuk (1983) distinguished four main varieties of West Polesian based on vocalism (Central Zaharoddzian, North-Eastern Zaharoddzian, Southern Zaharoddzian and Torokanian). It has been a common practice in Belarusian dialectology (and even between some interested West Polesian speakers) to use a set of shibboleth words to identify which group someone's variety belongs to. Amongst the most common diagnostic words are: *kinj* - *knj* - *kunj* - *kwonj* - *konj* 'horse'; *vin* - *vm* - *vun* - *vwon* - *von* '3SG.NOM.M pronoun; he'; *jabloko* - *jeblik(o)* - *jeblik(o)* 'apple'; and *pasok* - *pesok* - *pisok* - *pisok* 'sand'.

This description is mostly based on Central Zaharoddzian (the most widespread variety) and Torokanian (the less extensive variety, and the most peculiar in terms of vocalism). I also worked for a while in an area (Tolkovo, Drahičyn) where allegedly North Zaharoddzian was spoken, although I did not observe major differences with respect to Central Zaharoddzian. Moreover, in those places in which the Torokanian variety was spoken (primarily only spoken by elder people), I observed a notable tendency to switch to Central Zaharoddzian, sometimes more consciously (in the presence of non-locals) and sometimes less consciously. Fjodar D. Klimčuk (p.c.) also shared with me that in one of the few remnants of Torokanian varieties, Pare (Pinsk), he had observed that people were ‘bi-dialectal’ (besides people’s knowledge of Russian, Belarusian or Ukrainian), and that this is not a rare phenomenon to find in Western Polesie.

3.1.1. Phonemic inventory

In this subsection I present the phonemic inventory of West Polesian. However, for practical reasons (and aiming to make data easy to read to the linguist not familiar with Slavonic) I use adapted IPA representations to create a working orthography. As a result I omit most diacritics in my working orthography (i.e. dental and velarised articulation symbols, and I have a particular way of representing palatalisation which I explain in (§3.1.2.).

3.1.1.1. Consonants

According to my analysis, West Polesian has 23 consonants (summarised in Table 1), which distinguish from seven to eight places of articulation (since the only pharyngeal

is really a counterpart of the voiceless velar in its origins, as happens in BLM and ULM)³⁶ and seven modes of articulation.

Table 1 Consonant inventory of West Polesian

	Bilabial		Labio-dental		Dental		Alveolar		Palatal		Retroflex		Velar + pharyng.	
Stop	p	b			t̪	ɖ							k	
Fricative			f	v			s	z			ʃ	ʒ	x	ħ
Affricate				ts					ʃtʃ ^j	ɖʃ ^j	tʃ			
Nasal		m						n						
Lateral								lʲ						
Trill								r						
Glide			ʋ						j					

The articulation of /ɖʃ^j/ is problematic for representation as I have observed a significant amount of variation across varieties. Most frequently I represent it simply as /dj/.

I argue that /f/ should be included in the phonemic inventory. The emergence of /f/ as a phoneme, like in other Eastern Slavonic languages, is a late incursion, but in this case even later than in BLM or Russian. See some examples of sound correspondences of /f/ across Eastern Slavonic in Table 2.

³⁶ See a more detailed historical explanation in Shevelov (2015).

Table 2 Correspondences of /f/

Russian	Belarusian	West Polesian	Gloss
/fasolʲ/	/fasolʲ/	/pasolj/	‘bean’
/kartofli/	(The proper noun is /bulʲba/)	/kartopli/ or /kartoplja/	‘potatoes’
/afanasij/	/afanasij/	/apanasj/ or /panasi/	‘Afanasij’ (proper name, from <i>Athanasios</i>)
/fjodor/	/fjodar/	/xvedur/	‘Fedor’ (proper name from <i>Theodoron</i>)

Given that the amount of Russian loanwords with /f/ is considerable (e.g. /'fura/ ‘lorry, truck’; /fo'narık/ ‘torch’), I argue that it has entered the inventory.

3.1.1.2. Vowels

West Polesian has six vowels, represented in Figure 2 (from an impressionistic perspective). The vowel /ɪ/ ([ɪ]) matches the ULM <и> ([ɪ]) (transliterated as <y>) in quality, in opposition to the corresponding phoneme in BLM and Russian (<ы>, transliterated as <y>), which is a high central non-rounded vowel [ɪ]. This vowel is also lower than the corresponding vowel in Polish. Nevertheless, I have observed significant variation in the quality of /ɪ/, to which some speakers were sensitive, as a way of identifying the origin of the person in the area or even within the same village.

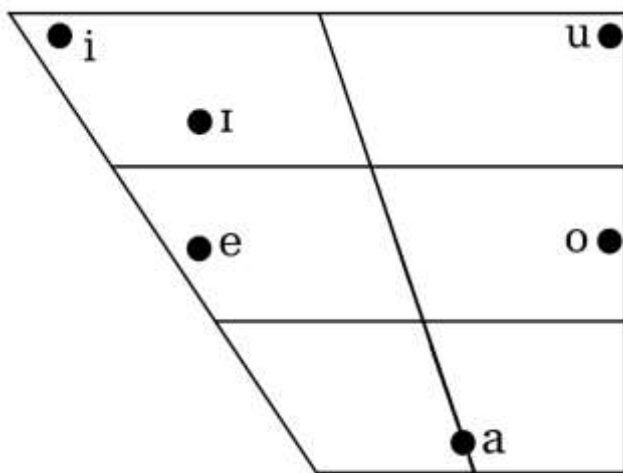


Figure 2 Impressionistic vowel inventory of West Polesian

All vowels can be headed and followed by the glide /j-/ (i.e. /aj/, /ej/, /ij/, /ɪj/, /oj/, /uj/, /ja/, /je/, /ji/, /jɪ/, /jo/, /ju/). All vowels can also appear with /-v/ in the coda (i.e. /av/, /ev/, /iv/, /ɪv/, /ov/, /uv/), although this last one appears mostly in inflectional suffixes.

3.1.2. Palatalisation

Palatalisation in West Polesian is a complex question which deserves a thesis on its own. I will briefly mention the problems and existing analyses and explain the decision I have adopted for my working transcription. First, Klimčuk's (1983) analysis of the old variant of Simonavičy (his hometown, Central Zaharoddzian) had six vowels, which distinguished /i/ and /ɪ/. However, his generation (born in the decade of 1930) had already an inventory of five vowels, and the quality of /i/ or /ɪ/ depends on the preceding consonant. Moreover, he proposes an extra level, which he calls 'palatalised' for every single consonant. Such an analysis is deeply rooted in the

Eastern Slavonic descriptive tradition.³⁷ Personally, I find such an analysis not to be very economic, although it must be said that Klimčuk provides evidence for all the pairs (from a corpus of more than 950 words). Second, in the light of the data I have gathered (primarily of Central Zaharoddzian and Torokanian varieties) it is not easy to assert that palatalisation creates a phonemic distinction between ‘soft’ (palatalised) and ‘hard’³⁸ (velarised, retroflex, or just simply non-palatalised) consonants, as in Russian and BLM. Moreover, the distinction between /i/ and /I/ is often unclear and changes considerably from one local variety to another. In fact, some authors (Hurski 1972, Lomtev 1956) have already challenged that distinction for Standard Ukrainian.³⁹ Third, it is a common practice in (Southern) Belarusian dialectology to distinguish properly ‘soft consonants’ C + <i> (i.e. palatals) vs ‘softened’ C + <и> (i.e. palatalised by accommodation) (Lena V. Levancèvič, p.c.). My approach for this work has been to classify palatalised consonants as the union of a consonant + /j/. This way the consonant accommodates to the articulation of the following one (i.e. there is regressive phonetic palatalisation), and I avoid proposing a complementary feature applying to every single consonant. As a result, /n/ has the allophones [n] when followed by /a/, /e/, /o/, /u/, and /I/ and [ɲ] when followed by /j/ or /i/ (see

³⁷ Yet, even in Russian, where the differences seem neater, there has been a long debate about the vowels /i/ and /i/ and palatalisation between schools, most notably the Muscovite School versus the St Petersburg School (opponent to the differences between vowels and palatalisation).

³⁸ Such terminology is commonly used to speak about the phonology of Eastern Slavonic languages. However, the term ‘hard’ is not very accurate from the perspective of the IPA conventions, as it can mean ‘velarised’, ‘retroflex’, or just simply ‘non-palatalised’.

³⁹ “Belarusian preserves the difference between the vowels <и> [i] and <ы> [ɨ]; cf. [...] byŭ [‘was’] and biŭ [‘hit’] [...] (in Ukrainian the vowels <и> and <ы> have merged into a single sound, which is between <и> and <ы> [...])”(Lomtev 1956: 17) [My translation].

§1.4.). In the context of this thesis, however, this can be understood as a mere graphic convention. Thus, if somebody chose to interpret any given C + /j/ cluster as a single palatalised phoneme, the content and conclusions of the thesis would not be altered at all.

3.1.3. Stress

Stress <'> is dynamic and phonemic across the entire lexicon in West Polesian. It often has a discriminatory function in the nominal paradigms (1) a. and a lot less frequently in verbal paradigms (1) b.:

West Polesian

- (1) a. (B6) [GEN SG] *ka'navi* vs. [NOM PL] *kana'vi* 'channel(s)'.
 b. (Tor1) [INF IPFV] *pozi'tfiti* vs. [INF PRF] *po'zɪtfiti* 'to borrow'

Since suffixation and stress are important for CASE/NUMBER marking but generally cannot be separated from each other, I will not segment nouns in the glosses.

3.1.4. Phonological and phonotactic rules in comparison with Eastern Slavonic

In this heading, I present a summary of the main phonological and phonotactic rules affecting West Polesian and Eastern Slavonic. The first three are shared by most Eastern Slavonic. More specifically, (§3.1.4.1.) *polnoglasie* is one of the distinctive marks of Eastern Slavonic and it is common to all; (§3.1.4.2.) final devoicing is also shared by all, although there is a debate around ULM; and (§3.1.4.3.) prothetic

consonants are also found in BLM and ULM, but in West Polesian this process appears to be more generalised. By contrast, the last point (§3.1.4.4.) covers those phonological phenomena in West Polesian that differ from the Eastern Slavonic group.

3.1.4.1. Polnoglasie (pleophony)

West Polesian has undergone the same process as the other Eastern Slavonic languages commonly known as ‘*polnoglasie*’ (a.k.a. “pleophony” or “full vocalization” (Sussex & Cubberley 2006: 36)).⁴⁰ That is to say, Common Slavonic consonant clusters of stop/fricative + liquid⁴¹ + mid-height vowel developed an epenthetic vowel between the two consonants in Eastern Slavonic (e.g. [Late Proto-Slavonic] **zôlto* > [WP, Russian] *zoloto* ‘gold’).

3.1.4.2. Final devoicing

As is common in Slavonic languages, all the final plosives undergo a process of devoicing when they appear in absolutely final position or before a stop. For example, the noun *xlib* ‘bread’ is pronounced [NOM.SG] /xlib/ > [xlip]; but [GEN.SG] /'xliba/ > ['xliba]. This final devoicing does not affect /v/ (as it does in BLM and Russian), because final /v/ and /l/ tend to transform into the glide /-ʋ/ (thus, inherently voiced).

⁴⁰ See for example the term ‘*polnoglasie*’ in Lomtev (1956: 15-16).

⁴¹ I.e. /l/ and /r-r/

3.1.4.3. Prothetic consonants

West Polesian has a preference for consonantal onsets. Many stems starting with a vowel take a prothetic /f/- or /v-/ depending on the variety. This also happens in BLM and ULM, which as a general rule take /v-/ in stressed (synchronically) /o-/ and /u-/ initial words. Yet, I have observed that this pattern is even more widespread than in BLM and ULM, affecting more vowels, as a result. According to Sussex & Cubberley (2006: 125-126) most of the words which undergo this process do so as a reflex from Common Slavonic /q/ (i.e. nasal o) after denasalisation (in contemporary Slavonic languages). See some examples of West Polesian prothetic vowels in Table 3:

Table 3 Sound correspondences of vowel initial words

West Polesian	BLM	Russian	Gloss
fiaurej	jaŭrej	evrej	‘Hebrew, Jew’
fiantosj	Anton /Antos’	Anton	‘Anton’ (proper name)
fiozero / vozero	vozera	ozero	‘lake’
fjutʃʲɪtsj(a) / vutʃʲɪtsj(a)	vučycca	učist’lja	‘to study, to learn’

3.1.4.4. Main phonological differences between Eastern Slavonic varieties

The main differences between BLM and WP are the absence of ‘*akanne*’, ‘*jakanne*’, ‘*dzekanne*’, ‘*cekanne*’, the realisation of the mid-front vowel and the phenomena around phonotactic restrictions and palatalisation. Starting from the vocalism, unstressed /o/ vowels in Contemporary Standard Russian (CSR) and BLM undergo

quality reduction, commonly known as ‘*akanne*’.⁴² For example the CSR verb *xodit’* is realised as [xΛ¹dʲitʲ] ‘to walk’, whereas its West Polesian cognate *xoditi* is realised as [xo¹diti]. Hence, West Polesian, like Standard Ukrainian (ULM), does not undergo *akanne*.

In addition to this, Belarusian (including most dialects) follows a phonological rule that distinguishes it from the rest of Slavonic varieties known as ‘*jakanne*’. According to this rule, every unstressed (especially in pretonic position)/e/ transforms into /ja/, unless the preceding consonant cannot be palatalised (e.g. /r/; [CSR] *re’bro*; [BLM] *ra’bro* ‘rib’). Nevertheless, this phonological rule does not apply to West Polesian.

Table 4 Eastern Slavonic sound correspondences and *jakanne*

Russian	Belarusian	Ukrainian	West Polesian	Gloss
se’lo	sja’lo	si’lo	si’lo	‘village’
te’per’	cja’per	te’per	ti’per	‘now’

In contrast, West Polesian and ULM have undergone a sound change affecting a phonological (vowel) system which is no longer there, known as ‘*ikanne*’ (in BLM).⁴³ Depending on the sub-variety of West Polesian, the vowel mutation has given /i/, /ɪ/ (‘*ykanne*’) or intermediate sounds between both,⁴⁴ whereas in ULM always /i/ (e.g.

⁴² The vowel reduction process is more complex in CSR than in BLM. In CSR only the /o/ or /a/ right before the stressed syllable reduces to [Λ], in all other unstressed positions /o/ and /a/ reduce to [ə].

⁴³ I have also documented ‘*okanne*’ very rarely, and even what Lena V. Levancèvič (p.c.) and Mackevič et al. (1964) call ‘*ukanne*’, in certain varieties of West Polesian. They refer to a mutation of the original vowel into /o/ (in the case of *okanne*) or /u/ (for *ukanne*); e.g. the infix *-ova-* transforms into *-uva-/uva-* in many varieties as in the verb *ros’kazuvati* ‘to tell [IPFV]’.

⁴⁴ See my concerns on palatalisation as a distinctive feature in (§3.1.2.).

[BLM, CSR] *xleb* [ULM] *xlib*; [WP] *xlib/xlīb* ‘bread’; or [BLM] *boh*; [CSR] *bog*; [ULM] *bih/boh*; [WP] *biĥ/bīĥ* ‘god’ (Hrinčenko 2010)). Sussex & Cubberley (2006) explain the origin of this process in Ukrainian along these lines:

“In Ukrainian, apart from /ě/, the other sources of the new high front /i/ were /o/ and /e/ in (new) closed syllables. This change is usually interpreted as the result of compensatory lengthening on this vowel for the loss of the *jer* in the next syllable. The interim stages were diphthongs of the [ɤɔ] type, seen in some texts and still in some dialects [...]. In addition to providing new sources of /i/, this development produced new alternations of /i/ and /o/ or /e/” (Sussex & Cubberley 2006:130).

Another phonological process very characteristic of all Belarusian varieties, but absent in West Polesian (and Eastern Slavonic), is the fricatisation of dental stops when followed by a palatal, commonly known as ‘*dzekanne*’ and ‘*cekanne*’ (dj > dzj; tj > tsj) (e.g. *dzjaŭčyna* ‘young girl’). This rule does not apply in West Polesian, and the speakers are very aware of this (in fact, whenever they want ‘to imitate Belarusians’ this is the first thing they do). Thus, West Polesian does not impose phonotactic restrictions (regarding palatalisation) on dental stops; consequently, the cognate of Belarusian *dzjaŭčyna* ‘young girl’ is *ˈdʲivtʲina* in West Polesian.

I have already mentioned the differences in the realisation of the Common Slavonic <ѣ> between Belarusian and Russian (as [i]) and Ukrainian and West Polesian (as [ɪ]). Now, in many contexts the Common Slavonic high-front vowel /i/ has become centralised in West Polesian. This means that it overlaps with the etymologically distinct CS /i/, and as a result, it still retains the possibility of being preceded by an initial glide (i.e. /jɪ/); as in *jɪstɪ* ‘eat’. Moreover, the centralisation of the high-front

vowel permits that vowel in the onset; e.g. *iti* ‘to go’. Conversely, no other Eastern Slavonic variety allows the combination of glides with such a vowel or have it in the onset (especially at the beginning of a word).⁴⁵ Nevertheless, it is important to emphasise that this is due to the historical evolution of the vowel /i/ > /ɪ/ in West Polesian.

In addition to this, Ukrainian and particularly West Polesian have a tendency to avoid palatalisation, which I suspect may not constitute a phonological contrast for some speakers (or at least not for almost every consonant pair as in Russian or Belarusian). For example, the Common Slavonic form of the infinitive suffixes was *-ti*. The vowel /i/, in general, caused palatalisation in Belarusian and Russian, but not in Ukrainian and West Polesian; e.g. [BLM] /kupitsʲ/; [CSR] /kupitʲ/; [WP] /kupɪtɪ/ ‘to buy’.

Another phonotactic restriction that affects Belarusian (and Belarusian dialects), but which does not affect the rest of the Eastern Slavonic family (including West Polesian) is that the trill /r/ cannot be combined with palatal vowels and diphthongs. For example, (Belarusian) /raʲka/; (Russian) /rʲeʲka/; (West Polesian) /rɪʲka/ or /rɪʲka/ ‘river’.

Finally, inflectional class-I nouns cannot bear stress on the PLURAL suffix in BLM and only marginally in CSR (when the suffix stress is in both SINGULAR and PLURAL).⁴⁶ However, this does not apply to Ukrainian and West Polesian; e.g. (BLM/RU) [NOM.SG] ʲpesnja, [NOM.PL] ʲpesni; (WP/ULM) [NOM.SG] pɪsnja, [NOM.PL] pɪsʲnɪ ‘songs’.

⁴⁵ No results were given in dictionaries of the respective languages of words starting with *y*-.

⁴⁶ Based on data from (Biryła 1986).

3.2. Morphology

In general, West Polesian morphology and syntax are not very different from the rest of the Eastern Slavonic family. Throughout this dissertation I will be narrowing the focus to those phenomena that make West Polesian different from the rest and which are of typological interest from a wider cross-linguistic perspective. Yet, before going into specific parameters, I provide a brief overview of the nominal morphology (§3.2.1.); verbal morphology (§3.2.2.); syntax (§3.3.); and ANIMACY (§3.4.), whilst drawing the attention to some of the particularities of West Polesian.

3.2.1. Nominal Morphology

As we would expect from any Eastern or Western Slavonic language, nominals inflect for CASE and NUMBER (as well as GENDER, when it comes to adjectives and other complements) which is realised in suffixes (fusionally). For example, the suffix *-um* as in (Tor1) *djtjum* ‘children [DAT PL]’ is a unique suffix that marks DATIVE CASE and PLURAL NUMBER at the same time.

Like the vast majority of Slavonic languages, West Polesian has three GENDERS: MASCULINE (M), FEMININE (F) and NEUTER (N). Even though there is a high degree of predictability of the inflectional classes based on GENDER, this is not always reliable, as happens with other Slavonic languages. The number of inflectional classes hardly differs from other members of the Slavonic family (particularly in the East), where there is an ongoing disagreement among Slavists on how to classify them. I do not

want to enter into that debate, so for practical reasons I present the following taxonomy, admitting that it is open for other analyses.

- Inflectional class I contains nouns ending in *-a*, which are in its majority FEMININE (e.g. *xata* (F) ‘house’, Table 5).
- Inflectional class II contains masculine nouns ending in a consonant (e.g. *sm* (M) ‘son’ in Table 5). Impressionistically, this seems the largest inflectional class, although inflectional class III and particularly class I are large as well.
- Inflectional class III contains nouns ending in *-o* and *-e* (e.g. *okno* (N) ‘window’ in Table 5). Inflectional class III holds many similarities with class II (in oblique cases, except for the GEN PL). For this reason, some prefer to treat them as two sub-classes of the same class. The vast majority of the nouns in this class are NEUTER.
- There is a small collection of oddities that do not fit into any of the other three classes which in the Slavonic tradition is often referred as “inflectional class IV”. Yet there are no unifying characteristics; hence, it is not accurate to say they conform an inflectional class on their own. This morphologically heterogeneous group contains nouns from every gender, although most of them are FEMININE (e.g. *mati* (F) ‘mother’ in Table 5).

Table 5 Paradigms of the inflectional classes in West Polesian

	<i>xata</i> (I) ‘house’		<i>sm</i> (II) ‘son’	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOM	¹ xat-a	xa ¹ t-i	sin	si ¹ n-i
ACC	¹ xat-u	xa ¹ t-i	¹ sin-a	si ¹ n-iv/sino ¹ v-ej
GEN	¹ xat-i	xa ¹ t-ej/ xat	¹ sin-a	si ¹ n-iv/sino ¹ v-ej
DAT	¹ xat-e	xa ¹ t-am	¹ sin-u/-ovi	si ¹ n-am
INS	¹ xat-oj	xa ¹ t-ami	¹ sin-om	si ¹ n-ami
LOC	¹ xat-e	xa ¹ t-ax	¹ sin-u/-e	si ¹ n-ax
	<i>okno</i> (III) ‘window’		<i>matr</i> (IV) ‘mother’	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOM	ok ¹ n-o	¹ vokn-a/ ¹ fiokn-i	¹ mat-i	mat ¹ r-i?
ACC	ok ¹ n-o	¹ vokn-a/ ¹ fiokn-i	¹ mat-i	mate ¹ r-ej
GEN	ok ¹ n-a	¹ vikon/ ¹ fiokon	¹ mat-eri	mate ¹ r-ej
DAT	ok ¹ n-i/-u	ok ¹ n-am	¹ mat-eri	mat ¹ rj-am
INS	ok ¹ n-om	ok ¹ n-ami	¹ matr-iju	mat ¹ rj-ami
LOC	ok ¹ n-i/-e	ok ¹ n-ax	¹ mat-eri	mat ¹ rj-ax

The number of CASE values is also not free from controversy. Corbett (2008, 2012) and Zaliznjak (1973, 2002) had already pointed out how problematic this question was for Russian, which is far better studied than West Polesian. According to Corbett (2012), the number of CASES in Russian could be between seven and ten. In West Polesian there are at least four uncontroversial cases: NOMINATIVE, GENITIVE, DATIVE and INSTRUMENTAL.

The ACCUSATIVE is not a very canonical CASE value.⁴⁷ As in many Slavonic languages, the ACCUSATIVE is only morphologically autonomous for the FEMININE SINGULAR adjectives and nouns from the inflectional class I in SINGULAR. Elsewhere the ACCUSATIVE is either syncretic with the NOMINATIVE, if the noun's referent is INANIMATE; or GENITIVE, if the noun's referent is ANIMATE (see more on ANIMACY in (§3.4.)). In spite of this, Corbett (2008) strongly argues for the legitimacy of the ACCUSATIVE as a core CASE value of the system.

West Polesian LOCATIVE is a non-canonical CASE value. Besides the fact that the LOCATIVE cannot stand on its own without a governing preposition (i.e. syntactically non-autonomous), the differences between LOCATIVE and DATIVE are often blurry. Inflectional class-I nouns do not distinguish the LOCATIVE SINGULAR from DATIVE SINGULAR (and in fact FEMININE adjectives and other NP constituents have a syncretic paradigm for OBLIQUE cases). Furthermore, inflectional class II and III nouns as well as MASCULINE and NEUTER adjectives very often use the DATIVE SINGULAR and LOCATIVE SINGULAR suffixes interchangeably:

(2) a. (T2.ov)

X**	priviz	minji	bibli-ju
X.	bring.PRF.PST.M.SG	1.SG.DAT	bible-ACC.SG

na	naʃ-im	jazık-ovı
in	POSS.1PL-LOC.SG	language-DAT.SG

‘X** brought me a Bible in our language.’

⁴⁷ See more on Canonical Typology and canonical CASE values in (§5.5.2.1.) and (§6.5.).

b. (Z7.1 00:24)

[...]	i	duma-je	treba	jix	potopi-ti	de-to
	and	think-3PL	necessary	3PL.ACC	dunk.PRF-INF	somewhere

u	njak-omu	bolot-ji,
in	some-DAT.SG	marsh-LOC.SG

u	njak-omu	tak-omu	poŋian-omu	mist-je
in	some-DAT.SG	kind_of-LDAT.SG	bad-DAT.SG	place-LOC.SG

‘[...] and [he] thinks he should dunk them perhaps in a marsh or some sort of dodgy place.’

In (2) a. the expected LOCATIVE SINGULAR form for inflectional class-II (-u or -e) is replaced by the form we would expect for the DATIVE SINGULAR -ovi. In (2) b. we would expect the LOCATIVE SINGULAR form of the NEUTER adjectives to be -om, but, it is -omu, instead, which is the form we would expect for the DATIVE SINGULAR NEUTER/MASCULINE.

Besides the core six, there are other possible CASE values in West Polesian, most of which match Corbett’s (2008, 2012) and Zaliznjak’s (1973, 2002) description of ‘secondary case values’ in Russian. Since there is already some literature on these, I only mention them briefly.

- The VOCATIVE: Even though it is missing in the varieties of a significant number of speakers (or it only exists for certain noun classes; e.g. FEMININE proper nouns) West Polesian has a VOCATIVE SINGULAR case. This case is peculiar because it can be realised by up to four (or five) different morphophonological

strategies (see more in (Roncero 2016) (e.g. [NOM SG] 'Xvedur > [VOC SG] 'Xvedure!; [NOM SG] 'Safa > [VOC SG] Sa'fo!).

- The ADNUMERATIVE. A special form that nouns take when governed by lower numerals (e.g. [NOM SG] *brat*; [2] '*bratr*'; [NOM PL] *bra'tr* 'brother(s)'). I do not consider the West Polesian ADNUMERATIVE to be a CASE value (not at least fully), but I will have a detailed analysis and discussion on this in (Chapter 5: Adnumerative forms).
- The SECOND GENITIVE. It is only available for nouns and, in principle, only for the SINGULAR sub-paradigm of inflectional class I and II nouns.⁴⁸ The SECOND GENITIVE has a partitive function, whilst the (normal) GENITIVE can have multiple functions. This case form is morphologically robust in BLM and well-defined (e.g. [SECOND GEN] *cement-u*; [GEN SG] *cement-a* 'cement'), whereas this form seems marginal in West Polesian (e.g. [SECOND GEN] *tsukr-u*; [GEN SG] ?*tsukr-i* 'sugar').

3.2.2. Verbal morphology

Synchronically, West Polesian verbal morphology distinguishes three PERSON values and two NUMBER values (see for example (Table 6)).⁴⁹ Common Slavonic had a DUAL number, which has survived eroded for nouns in many contemporary Slavonic languages, but has for the most been lost in verbal morphology (with the exception of Slovene and Sorbian).

⁴⁸ See the discussion on the GREATER ADNUMERATIVE /SECOND GENITIVE (PLURAL) in (§5.3.).

⁴⁹ Although this is arguable when it comes to adnumerative forms, but I discuss this later in (Chapter 5).

Table 6 PRESENT TENSE paradigm of the verb *iti* ‘to go (on foot)’ in WP⁵⁰

1SG	idu	1PL	idam/idem(o)
2SG	idaʃ/ideʃ	2PL	idita/idite
3SG	ida/ide	3PL	idutj

Like all the Eastern Slavonic varieties, West Polesian distinguishes three TENSE values: PAST - PRESENT - FUTURE; although for Sussex & Cubberley (2006: 243) the division is rather PAST vs NON-PAST.⁵¹ The vast majority of PAST forms are marked by the additional suffix *-l* (which transforms into *-v* in absolute final position), in comparison to the non-past forms.⁵² In addition to this, the PRESENT and the FUTURE forms mark PERSON and NUMBER, the system changes radically with the PAST tense values. In (3) a. the verb *batʃiti* ‘to see’ is inflected for 1SG in the PRESENT tense (the unmarked form), whereas in (3) b. the verb also marks PERFECTIVE aspect, which transforms it into a PERFECTIVE FUTURE tense form (§7.4.6.).

- (3) a. zare ja **batʃ-u** 'brat-a
 now 1.NOM.SG see-1.SG brother-ACC.SG
 ‘Now I see [my] brother.’
- b. zavtra ja **pobatʃ-u** 'brat-a
 tomorrow 1.NOM.SG see.PERF-1.SG brother-ACC.SG.M
 ‘Tomorrow I will see [my] brother.’

⁵⁰ Given the different types of vocalism co-existing, in these examples I give the two possible forms of this paradigm across varieties. In the majority of villages I worked, speakers used the *-/e/* paradigm (at least in Central Zaharoddzian), *-/a/* being rather a minority (mostly Torokanian). Yet as previously said, speakers of Torokanian varieties often code-switch with Central Zaharoddzian (at least vocalism).

⁵¹ And there is a big debate on whether the FUTURE can be considered a TENSE value at all in Slavonic, which I explain in further detail in (§7.1.2.).

⁵² A few PAST tense stems have suppletive forms, such as [INF IPFV] *iti*; but [IPFV PST M SG] *foʋ* ‘go’.

In the PAST tense WP verbs (as well as the rest of Eastern Slavonic varieties) agree in GENDER with their head, for the SINGULAR (i.e. they only distinguish GENDER and NUMBER, but not PERSON) and only in NUMBER in the PLURAL.⁵³ Picking up the last example, the verb *'batʃiti* ‘to see’ has four forms in the PAST tense, three for the SINGULAR, and all the PLURAL forms are syncretic (4).

	SINGULAR		PLURAL	
(4) a.	vin	¹ batʃɪ- v	d.	vonɪ ¹ batʃɪ- l-ɪ
	3SG.NOM.M	see-PST.M.SG		3PL.NOM see-PST-PL
	‘He saw.’			‘They saw.’
b.	vona	¹ batʃɪ- l-a	e.	mi ¹ batʃɪ- l-ɪ
	3SG.NOM.F	see-PST-F.SG		1PL.NOM see-PST-PL
	‘She saw.’			‘We saw.’
c.	vono	¹ batʃɪ- l-o		
	3SG.NOM.N	see-PST-N.SG		
	‘It [NEUTER] saw.’			

When it comes to ASPECT, West Polesian distinguishes between PERFECTIVE and IMPERFECTIVE verbs like the rest of the Slavonic family.⁵⁴ IMPERFECTIVE verbal bases are most frequently the default form, from which PERFECTIVE forms are derived. Yet, when these are secondarily derived from PERFECTIVES, the IMPERFECTIVES are marked by the infix *-iva/uva* (e.g. [PRF] *roskazati* > [IPFV] *roskazuvati* ‘to tell’). The rules for derivation of the PERFECTIVE verbal bases are unpredictable. Most frequently they are

⁵³ Note that the failure to distinguish GENDER in the PLURAL is a property of the agreement system as a whole, not a property of PAST tense verbs.

⁵⁴ “[W]hich are not to be confused with the perfect and imperfect tenses” (Sussex & Cubberley 2006: 244).

derived by prefixation (e.g. [IPFV] *jixatɪ* [PRF] *pojixatɪ* ‘to go (by means of transport)’); and very rarely by vowel or stress shift (as in (2) b.).

The IMPERFECTIVE verbal base is the source for the PRESENT tense, the IMPERFECTIVE PAST tense and most of the PERIPHRASTIC FUTURE tense forms, which I describe in (Chapter 7).⁵⁵ The PERFECTIVE verbal base gives the PERFECTIVE PAST tense, the PERFECTIVE PRESENT (or FUTURE) and usually the DE-OBLIGATIVE (periphrastic) future (which I describe in greater detail in (Chapter 7)). Note that the PRESENT tense and the IMPERFECTIVE aspect are the unmarked form for the majority of verbs. For this reason, I only specify them in the glosses when they are overtly marked.

Reflexive verbs are marked by the suffix *-sj(a)* after the TENSE/PERSON/NUMBER/GENDER suffix; e.g. [INFINITIVE] *ɦutɦɨsj(a)*; [PRS.1SG] *ɦutɦusj(a)*; [IPFV.PST.M.SG] *ɦutɦɨvsja* ‘to study’. Note that I have documented multiple times the same speaker using the same verb (with same PERSON/NUMBER inflection) produced with and without the final *-a*. I have not been able to find any clear phonological rules to predict this. Hence, I believe there is free variation and I transcribe it each time the way the speakers produced it. When a speaker produced the same form but with two different realisations of the reflexive suffix I put the final *(-a)* in brackets.

3.3. Syntax

Like the vast majority of European languages, West Polesian has NOMINATIVE-ACCUSATIVE alignment. The unmarked word order is SVO as (5) a., nevertheless, there

⁵⁵ N.B.: Most Slavonic languages only have one periphrastic future construction.

is relative freedom in word order as in the rest of the Eastern Slavonic family (particularly in informal speech), thus sentences such as (5) b. (i.e. SOV) are also acceptable and not rare in the corpus.

(5) a. (P4.1.1 00:55)

muŕtŕ:in-I	tŕita-l-I	fi-etu	knifi-u,	tserkovn-uju
man-NOM.PL	read-PST-PL	that-ACC.SG.F	book(F)-ACC.SG	churchly-ACC.SG.F

‘Men used to read that church book.’

b. (Tor 1.4 02:24)

vin	vs-ix	kur-ej	pokra-u
3SG.NOM.M	all-ACC.PL	chicken-ACC = GEN.PL	steal.PRF-PST-M.SG

‘He stole all the chicken.’

In West Polesian the most usual word order for possessors is after the noun (particularly when referring to kinship and relationship terms) (as in (5) a.). In some idiolects this ‘uncommon’ (for Eastern Slavonic) word order allows the insertion of other constituents between the possessor and the noun as in (6) a-b.

(6) a. (T3.10 00:29)

ŕtŕe	odin	utŕitil	ŕiv-e	m-ij
still	one. NOM.SG.M	teacher.NOM.SG	live-3SG	POSS.1SG.NOM.SG

‘One of my teachers is still alive.’

b. (B6.2 3:37)

bo jak	batjk-o	prif-ov	m-ij	z	ŕermanj-ii	[..]
as COMP	father-NOM.SG	arrive.PRF-PST.M.SG	POSS.1SG.NOM.SG	from	Germany-GEN.SG	

‘So when my father arrived from Germany [...].’

c. (Z2.4a 01:16)

veljmo	ljubi-l-a	muž-a	svojfo
very	love-PST-F.SG	husband(M).ACC.SG.M	own-ACC.SG.M

‘I loved my husband a lot.’

Quantification also has special morphosyntactic effects, but I will dedicate (Chapters 4-6) to this specific question.

3.4. ANIMACY

ANIMACY has been one of the most studied parameters in typology in the last decades since Smith-Stark’s (1974) study. ANIMACY has been also quite widely studied in the Slavonic family (Eckhoff 2015, Huntley 1980, Vaillant 1958). Some authors, such as Corbett (1991) have considered it a sub-gender, and there seem to be good reasons to second this idea, since ANIMACY is mainly observable for its effects on syntax.

The traditional departure point or the main object of study for ANIMACY in Slavonic has been the behaviour of the ACCUSATIVE (for class II and III nouns and and/or agreement targets of MASCULINE nouns in the SINGULAR; and for any noun from any inflection class in the PLURAL).⁵⁶ The ACCUSATIVE is a non-autonomous CASE value (except for the SINGULAR sub-paradigm of inflectional class I), which in the relevant inflectional classes (and GENDERS) takes either the form of the NOMINATIVE if the referent is INANIMATE, which is often just described as [ACC = NOM]. Conversely, it

⁵⁶ This excludes the distinction between [VIRILE] and [NON-VIRILE] in languages like Polish, where it also affects the agreement of the PAST tense and some of the FUTURE tense PLURAL inflection. However, the distinction of [VIRILE] and [NON-VIRILE] seems a step further in the separation between [ANIMATE] and [INANIMATE] which West Polesian has not developed.

takes the form of the GENITIVE if the noun's referent is ANIMATE (usually presented as [ACC = GEN] in the literature).⁵⁷ In the example bellow of Russian, (7) a. has an ANIMATE noun in ACCUSATIVE, (7) b. an INANIMATE noun in ACCUSATIVE and (7) c. has the GENITIVE form of the same INANIMATE noun:

Russian

- (7) a. Ja viž-u mal'čik-a
 1SG.NOM see-1SG boy-ACC = GEN.SG
 'I see [the/a] boy.'
- b. Ja viž-u kamen'
 1SG.NOM see-1SG stone.ACC = NOM.SG
 'I see [the/a] stone.'
- c. Vod-a vytek-l-a iz kamn-ja
 water-NOM.SG flow_out.PRF-PST-FEM.SG from stone-GEN.SG
 'Water came out of [the/a] stone'.

However, ANIMACY in West Polesian has more shades of grey than in Russian. First, it had been already vaguely mentioned by some authors (Huntley 1980, Mackevič et al. 1964) (including myself (Roncero 2015)) that in Southwestern Belarusian nouns denoting animals are INANIMATE in the PLURAL sub-paradigm. After undertaking fieldwork on West Polesian, I can confirm that in most varieties (perhaps slightly more

⁵⁷ Corbett (2011) says on animacy in Russian that "[o]n the one hand, the animacy distinction is severely limited in that it is found within just one case (and it is non-canonical in this respect). On the other hand it is a central part of the system, affecting nouns, pronouns, almost all adjectives (those that can occurs in attributive function) and some numerals" (Corbett 2011: 455).

remarkably amongst older speakers) nouns denoting animals inflect like INANIMATES in the ACCUSATIVE PLURAL. Nevertheless, many speakers can on certain occasions also treat these nouns like ANIMATES (see (8) b-c.) produced by the same speaker (Z4)), that is to say, there is some fluidity in this ANIMACY (cf. (8) a-c.)):

(8) a. (B6.1 15:30)

koro'vi	derʒa-l-i,	'svinje	derʒa-l-i
COW.ACC = NOM.PL	own-PST-PL	pig.ACC = NOM.PL	own-PST-PL

‘We owned cows, we owned pigs.’

b. (Z4.2 08:51)

dofljada-je	koro'vi
look_after-3SG	COW.ACC=NOM.PL

‘[She] looks after cows.’

c. (Z4.5 00:56)

tak,	svi'nej	kormi-l-i	sirovotk-eju
yes	pig.ACC = GEN.PL	feed-PST-PL	whey-INS.SG

‘We used to feed the pigs with whey.’

That is to say, for most speakers there is a lexico-syntactic split between the SINGULAR (where ANIMACY is restricted to MASCULINE inflection class-IIa nouns) and PLURAL sub-paradigms for nouns denoting animals.

Besides the effects of ANIMACY in the choice of syncretism for the ACCUSATIVE, it has an effect in other areas of the morphosyntax. First, only human animates can have a VOCATIVE form. T2 had preserved very well the VOCATIVE SINGULAR FEMININE (e.g. *Babo hanno!* ‘Oh, granny Hanna!’) and he had given female names to his goats (e.g. *hruša*

‘pear’) which he would often talk to. Nevertheless, in all the six weeks I spent with him, I never heard him using a single VOCATIVE form with the animals. Second, impressionistically only ANIMATE (probably only humans) can appear before the possessor (e.g. (T1) *matɪ jɪfɒ/tofɒ xloptsɑ* ‘his mother/ that boy’s mother’; *okna toj xatɪ firaznɪja* but **okna jɪjɪ firaznɪja* ‘the windows of that house are dirty’). However, I need more evidence to confirm this. Third, pronominal numerals (§4.4.) can only be used with humans. Conversely, collective numerals can be used with both humans and animals in the NOMINATIVE, but in the majority of contexts only with humans in the ACCUSATIVE (§4.3.2.). I expand more on the different classes of numerals in the next chapter (Chapter 4).

In conclusion, West Polesian needs an additional category (in comparison to the Eastern Slavonic system) for ANIMACY to give account of the data. Polish has a similar three-way division of the Animacy Hierarchy, although the division is not identical. The top category in Polish is VIRILE (thus INANIMATE - ANIMATE (MASCULINE) - VIRILE), whereas in West Polesian is more general, simply HUMAN (thus, INANIMATE - ANIMATE-HUMAN). Furthermore, ANIMACY in West Polesian is a parameter that is beyond the ACCUSATIVE cells and nominal agreement, as we are going to see in this work.

3.5. Summary

In sum, West Polesian shares in common the core features of the Eastern Slavonic group. In the area of phonology pleophony, distinctive dynamic stress and final devoicing. In the area of morphology, CASE/NUMBER inflection for nouns (with almost

identical inflectional classes), and a TENSE split in the marking of PERSON/NUMBER/GENDER between PAST and NON-PAST verbal forms. And in respect of syntax and ANIMACY, the SVO word order and the non-autonomy of the ACCUSATIVE (except for infl. class I FEMININE nouns). When it comes to the distinctive features of West Polesian within the Eastern Slavonic group, we have seen a few. In the area of phonology the peculiarity of palatalisation; the quality of the vowel /ɪ/ and fewer phonotactic restrictions for this vowel due the centralisation of CES /i/ in many contexts (*ykanne*). In the area of morphology, so far we should mention the multiple markings of the VOCATIVE and the ongoing process of syncretism between the DATIVE and the LOCATIVE. In syntax, the most distinctive feature is that possessors tend to appear postnominally. Finally, concerning ANIMACY, West Polesian stands out from the rest of the Eastern Slavonic family, thanks to the development of a third (intermediate) category for NON-HUMAN ANIMATE nouns (i.e. farm animals), yet it is typologically very close to the Polish three-way distinction.

Part II. Main research outcomes

Section I. Numeral phrases and quantification

Chapter 4

Numeral phrases and numerals in West Polesian

In Slavonic languages, the field of numeral phrases and quantification is complex and, at the same time, fascinating. As a result, this has attracted the attention of many linguists over the decades (Akiner 1983, Babby 1987, Corbett 1978a, 1978b, 1978c, 1983, Franks 1995, Hurski 1972, Kim 2009, Mel'čuk 1985, Pereltsvaig 2009, Rappaport 2002, Suprun 1961, Viellard 2011, Vjarxoŭ 1961, Žolobov 2003, 2007), just to mention a few. In fact, Viellard (2011) points out that the first descriptive work of numerals in Slavonic came from Fr. Dobrovský (1822) at the beginning of the nineteenth century. However, when it comes to West Polesian, the system of numeral phrases and quantification is even more complex, yet unstudied. For this reason, more than two thirds of the core content of my thesis deal with all of the phenomena surrounding numeral phrases and quantification in West Polesian.

In this chapter, after introducing some properties of numeral phrases in West Polesian (§4.1.), I devote most of my attention to three types of numerals in West Polesian: (§4.2.) **cardinals** (*tŕtŕi* 'four'); (§4.3.) **collectives** (e.g. *troje* 'the three together'); and (§4.4.) **pronominals** (*dvox* 'the two of [us]'), among which, the first ones, the cardinals are the most prominent. With cardinal numerals and certain quantifiers, a small group of nouns display suppletive patterns in West Polesian. However, suppletion is treated separately in (Chapter 6). I continue in this chapter with a brief presentation of other types of numerals (§4.5.): (§4.5.1.) ordinals (e.g. *druŕi* 'second'); (§4.5.2.) fractions (e.g. *pov* 'half'); (§4.5.3.) distributives (e.g. *oboje* 'both'); and

(§4.5.4.) quasi-adverbial numerals (e.g. *vdvox* ‘both together’). Finally, (§4.6.) I close with a summary of the main phenomena and conclusions.

I should make clear from the beginning that I treat the types of numerals and **NUMBER** (or **NUMBER** values) as completely different terms. The first one refers to the possibilities or ways quantification of items can take place in a particular language. The second one refers to the possible values that can be found in a language; that is to say, how the individuation of items is codified (e.g. **SINGULAR**, **PLURAL**, etc.). I discuss this further in (Chapter 5), but so far, the most complete theoretical work on this topic can be found in Corbett (2000).

4.1. General overview of numerals in West Polesian

The system of numerals in Slavonic languages is known for being morphologically rich and uncommon from the point of view of typology. Vaillant (1958) points out that:

“[t]he system of Indo-European was certainly not simple, and it is partly obscure, even for the names of the decenes. Slavonic has created a new system that is logical, but complex when it comes to inflection and the agreement of nouns that go with the numerals” (Vaillant 1958: 650) [My translation].

West Polesian has different classes of numerals which are present in most (if not all) Slavonic languages. These include (§4.2.) cardinals (*tri* ‘three’); (§4.3.) collectives (e.g. *troje* ‘the three together’); (§4.5.1.) ordinals (e.g. *treći* ‘third’); and (§4.5.2.) fractions (e.g. *povtora*

‘one and a half’) (see for example Vjarxoŭ’s (1961) work on numerals in Belarusian and Eastern Slavonic and his classification). But West Polesian has other types of numerals, which are less widespread or even unique, in the case of pronominal numerals. These include (§4.4.) pronominals (*trox* ‘the three of [us]’); (§4.5.3.) distributives (e.g. *oboje* ‘both’); and (§4.5.4.) quasi-adverbial numerals (e.g. *vtrox* ‘all three together’).

According to Kim (2009) there is a big debate on whether numerals can be heads or not (particularly in Slavonic).

“With regard to the headedness of Russian numeral phrases, there exists no consensus among linguists, although there is a prevailing opinion that numerals are not heads, at least in oblique cases (Babby 1987; Neidle 1988; Peškovskij 2001)” (Kim 2009: 9. See references there).

Nevertheless, Kim (2009) also presents evidence from other authors (including Corbett (1993)⁵⁸ and Rappaport (2002, 2003)) who favour a second view; i.e. that numerals are more likely to be governors:

“Rappaport (2002, 2003) maintains that nouns are heads of Russian numeral phrases by highlighting the fact that nouns have inherent value for φ -features, while numerals and adjectives cannot have an inherent value for φ -features” (Kim 2009: 22).

I will be giving further details on the government and agreement of numerals in the following sections and especially in the next chapter (§5.2.3.). Nevertheless, I already advance that I have also adopted a similar position, which is partly challenged in (§5.4.).

⁵⁸ However, Corbett (1983: 216) suggests the opposite “[...] it is indeed the noun which is the head of quantified expressions [with higher numerals], and so is the agreement controller”.

In short, I believe that numerals should be treated as heads in West Polesian. The reason behind this is that all the numerals greater than ‘one’ (including cardinals, collectives, distributives, fractions and pronominals) trigger similar (and most often the same) agreement structures:

- They trigger PLURAL (most often 3PL) verbal agreement in the PRESENT tense.
- They trigger NEUTER SINGULAR agreement in the PAST tense (although with cardinal numerals PLURAL agreement is sometimes also available, being even obligatory for a few verbs).
- Most often the adjectives inside the numeral phrase (NumP) take GENITIVE PLURAL.

For the rest, ordinal numerals are to be treated as mere adjectives and, for this reason, they cannot trigger any type of agreement.

4.2. Cardinal numerals

In this section, I give a general overview of West Polesian cardinal numerals. In general, morphologically, they differ very little from the rest of the Slavonic languages, but morphosyntactically they have a peculiar behaviour which I describe in the next chapter. In this section, firstly, (§4.2.1.) I present the main cardinal numerals, offer some remarks about their morphology, and classify them into groups. Secondly, (§4.2.2.) I give a glimpse of the general morphosyntactic behaviour of cardinal numerals and what they trigger in numeral phrases (NumPs), and this point helps understanding the last one, (§4.2.3.) postnominal numerals, which mark an approximate quantity.

4.2.1. General morphological remarks

The West Polesian counting base is ‘ten’, although, as in many other Indo-European languages (e.g. Portuguese, English), the numbers up to ‘twenty’ have a special form. Numbers from ‘eleven’ to ‘nineteen’ have a different syntactic structure from any other numeral in West Polesian, since the unit precedes the decimal (as in German). The rest of the forms up to *sto* ‘a hundred’ are derived from ‘one’ to ‘ten’. Thus, they reproduce the pattern of ‘one’, lower numerals (‘two’ to ‘four’) and higher numerals (‘five’ to ‘twenty’) (the decimals behaving like ‘ten’). I summarise the main cardinal numerals in West Polesian in Table 6.

From a purely morphological perspective, West Polesian cardinal numerals do not differ from other Eastern Slavonic numerals. Following the common practice in Slavonic typology (based on morphosyntactic agreement), I will classify numerals into three groups ‘one’ (§4.2.1.1.); lower numerals (‘two’ to ‘four’) (§4.2.1.2.); and higher numerals (‘five’ to ‘twenty’; hundreds, thousands, etc.) (§4.2.1.3.). For the premises of this work, I will focus on numerals up to ‘twenty’, particularly the first five decimals.

Table 6 Main cardinal numerals in West Polesian

(including phonological variations)

Numeral	Gloss	Numeral	Gloss
<i>odm</i> (M) <i>odna</i> (F) <i>odno</i> (N) <i>odnu</i> (PL)	‘one’	<i>'dvatsitj odm</i> (M) <i>odna</i> (F) <i>odno</i> (N)	‘twenty-one’
<i>dva</i> (M/N) <i>dvi/I</i> (F)	‘two’	<i>'dvatsitj dva</i> (M/N) <i>dvi/I</i> (F)	‘twenty-two’
<i>tri</i>	‘three’	<i>'tritsetj</i>	‘thirty’
<i>'tʃtʉrɪ/ 'tʃʰtʉrɪ/ tʃotʉrɪ</i> (rare)	‘four’	<i>'sorok</i>	‘forty’
<i>pjetj/ pjatj</i>	‘five’	<i>pitjdɪ'sjat</i>	‘fifty’
<i>ʃest(j)/ ʃist(j)</i>	‘six’	<i>ʃistdɪ'sjat</i>	‘sixty’
<i>sim</i>	‘seven’	<i>simdɪ'sjat</i>	‘seventy’
<i>'vosim/ 'visim</i>	‘eight’	<i>'visimdisjat</i>	‘eighty’
<i>'devet(j)/ 'devit(j)</i>	‘nine’	<i>dviɪ'nosta</i>	‘ninety’
<i>'desitj</i>	‘ten’	<i>sto</i>	‘one hundred’
<i>'odmatsitj/ odi'natsitj</i>	‘eleven’	<i>'dvjestɪ</i>	‘two hundred’
<i>dvi'natsitj</i>	‘twelve’	<i>'trista</i>	‘three hundred’
<i>tri'natsitj</i>	‘thirteen’	<i>tʃtʉr'rista</i>	‘four hundred’
<i>'tʃtʉrnatsitj/ tʃtʉr'natsitj</i>	‘fourteen’	<i>pitj'sot</i>	‘five hundred’
<i>pitj'natsitj</i>	‘fifteen’	<i>ʃistj'sot</i>	‘six hundred’
<i>ʃistj'natsitj</i>	‘sixteen’	<i>sim'sot</i>	‘seven hundred’
<i>sim'natsitj</i>	‘seventeen’	<i>visim'sot</i>	‘eight hundred’
<i>vosim'natsitj/ visim'natsitj?</i>	‘eighteen’	<i>dviɪtj'sot</i>	‘nine hundred’
<i>dviɪ'natsitj</i>	‘nineteen’	<i>'tʉsɔtʃa</i>	‘one thousand’
<i>'dvatsitj</i>	‘twenty’	<i>milɔn</i>	‘one milion’

4.2.1.1. The numeral ‘one’

As in other Slavonic languages (Corbett 1978b), the numeral ‘one’ is the numeral which is closest to adjectives. It agrees in GENDER and NUMBER with its target.

(9) a.	odn-a	¹ dʲɪrʊtʃm-a	b.	odn-e	drʲtʲj-o
	one-NOM.SG.F	girl(F)-NOM.SG		one-NOM.SG.N	child(N)-NOM.SG
	‘One girl.’			‘One child.’	
c.	odn	¹ xlopets	d.	odn-i	sanʲk-i
	one.NOM.SG.M	boy(M).NOM.SG		one-NOM.PL	sledge-NOM.PL
	‘One boy.’			‘One sledge’ ⁵⁹	

4.2.1.2. Lower numerals

The group is formed by ‘two’, ‘three’ and ‘four’ (and derived forms). Inside this group, there are significant differences between ‘two’ and the rest, which creates a sub-group of its own.

4.2.1.2.1. The numeral ‘two’

The numeral ‘two’ preserves its adjectival properties, but with more limitations than ‘one’. It is inherently NON-SINGULAR, and only distinguishes FEMININE from NON-FEMININE (i.e. MASCULINE and NEUTER). In any case, the GENDER distinction is only relevant for the NOMINATIVE, and in some varieties the ACCUSATIVE as well. It shares this particularly with other Slavonic languages and it is well attested in the literature (see Akiner 1983,

⁵⁹ NB: Besides its pure quantificational meaning, *odn* is often employed with an adverbial function meaning ‘just, only’, especially when it is used in plural. For example (T3.8 00:01) *ja a[d]no zare derzu křiʃi i muʃi, a bilʃ u mene numa* ‘Now I just own mice and rats, I don’t own anything else’.

Corbett 1978b, Kim 2009, Mayo 1976, Pugh & Press 1999, Shevelov 1963, Vjarxoŭ 1961, just to mention a few).

- | | | | | | |
|---------|--------------|------------------------|----|-----------------|---------------|
| (10) a. | dv- <i>l</i> | ¹ dʒɪʋtʃɪnɪ | b. | dv- <i>a</i> | dɪ'tʃatɪ |
| | two-NOM.F | girl(F).ADNM | | two-NOM.N | child(N).ADNM |
| | ‘Two girls.’ | | | ‘Two children.’ | |

4.2.1.2.2. The numerals ‘three’ and ‘four’

The numerals ‘three’ (*tri*) and ‘four’ (*tʃtɪrɪ*) do not agree in GENDER or NUMBER with their target, but as with the rest of the cardinal numerals they agree in CASE in non-direct cases (e.g. neither NOMINATIVE, nor ACCUSATIVE)⁶⁰ (see Footnote 61).

- | | | | | | |
|---------|----------------|------------------------|----|-------------------|---------------|
| (11) a. | tri | ¹ dʒɪʋtʃɪnɪ | b. | tri | dɪ'tʃatɪ |
| | three.NOM | girl(F).ADNM | | three.NOM | child(N).ADNM |
| | ‘Three girls.’ | | | ‘Three children.’ | |
-
- | | | | | | |
|----|-----------------------|--------------------------|----|------------------------|-----------------|
| c. | dv-om | ¹ dʒɪʋtʃɪn-am | d. | tr-ima | dɪ'tʃ-mɪ |
| | two-DAT | girl(F)-DAT.PL | | three-INS | child(N)-INS.PL |
| | ‘To (the) two girls.’ | | | ‘With three children.’ | |

⁶⁰ Lower numerals also agree in CASE with the target if we analyse the ADNUMERATIVE form as a NUMBER value. However, I will develop this idea extensively in the next chapter (§5.5).

4.2.1.3. Higher numerals

4.2.1.3.1. Basic higher numerals

These include numerals from ‘five’ to ‘twenty’ (and derived forms; e.g. *dvatsit pjetj* ‘twenty-five’). In the same vein as ‘three’ (*tri*) and ‘four’ (*tŕiri*), they are also invariable in GENDER.

- | | | | | | |
|---------|---------------|----------------|----|------------------|-------------------|
| (12) a. | pjetj | di'vok | b. | pjetj | bra't-iv |
| | five.NOM | girl(F).GEN.PL | | five.NOM | brother(M)-GEN.PL |
| | ‘Five girls.’ | | | ‘Five brothers.’ | |

They also agree in CASE with their targets in NON-DIRECT CASES, although CASE-agreement decreases its rigorousness the higher the numeral (13) a.^{II}, b.^{II}.⁶¹

- | | | | | | |
|----------------------|---------------------------|-------------------|------------------|-----------|-------------------|
| (13) a. ^I | pŕit-i | diu'k-am | a. ^{II} | pjetj | diu'k-am |
| | five-DAT | girl(F)-DAT.PL | | five.NOM? | girl(F)-DAT.PL |
| | ‘To (the) five girls.’ | | | | |
| b. ^I | pŕit-i | bra't-am | b. ^{II} | pjetj | bra't-am |
| | five-DAT | brother(M)-DAT.PL | | five.NOM? | brother(M)-DAT.PL |
| | ‘To (the) five brothers.’ | | | | |

⁶¹ Although I need more evidence to assert this for West Polesian, the higher the numeral (starting from ‘five’), the greater the flexibility to omit the agreement in NON-DIRECT CASES, as Corbett (1983, 2000) pointed out happens in other Slavonic languages.

4.2.1.3.2. Hundreds, thousands and millions

This could be analysed as a subgroup within higher numerals. They follow the general syntactic rule of placing bigger units before the smaller units (except for those instances in which the smaller unit is determining the bigger one. For example, in (14) *dva* ‘two’ is determining *milioni* ‘millions’.

- (14) (T1.6 04:31)
- | | | | | | |
|--------|-------------|----------------|------|-----|----------|
| krome | dv-a | mil'oni | bilʃ | ni | daj-utj |
| except | two-NOM.M | million.ADNM | more | NEG | give-3PL |
- ‘They don’t give more than two million.’

In DIRECT CASES (especially in NOMINATIVE) they behave like higher numerals, but as a general rule they do not agree in CASE. Nevertheless, as far as I have been able to observe this subgroup presents no innovations or oddities with respect to the Eastern Slavonic group, and thus I do not deal with it in the present work.

4.2.2. Morphosyntactic behaviour

For this work, I concentrate exclusively on the above-described groups of lower and higher numerals (up to ‘twenty’). I will not enter into more details about morphosyntactic agreement and government relations between cardinal numerals and NPs, as I deal extensively with this in the next chapter (Chapter 5). However, in order to understand postnominal numerals (§4.2.3.), I need to give a very concise overview of this.

Cardinal numerals, in their most basic use, quantify a (precise) number of items such as *tri stra'mni* ‘three ladders’. When they are fulfilling this function, they must appear before the NP they are quantifying.

When numerals higher than ‘one’ precede an NP, the agreement becomes particularly tricky, and this is one of the primary questions I try to explore in this work. When a lower numeral is heading the NP, the noun takes ADNUMERATIVE (henceforth, ADNM) (with the possibility of alternatively adopting forms that resemble GEN SG and NOM PL, at least at the surface level). Adjectives and other complements can also take either GEN PL or NOM PL (at least in the surface level). Higher numerals cause the noun and the rest of the constituents within the NP to take GEN PL.⁶² Both lower and higher numerals trigger 3PL verbal agreement in the PRESENT tense, but (most often) NEUTER SINGULAR in the PAST tense.

4.2.3. Postnominal numerals

Numerals, particularly cardinal numerals, can appear after a noun in order to denote an approximate quantity. Compare (15) a. cardinal numeral, vs. (15) b., a postnominal numeral.

- | | |
|---|--|
| (15) a. pjetj tʃolo'vik
five.NOM? person(M).GEN.PL
‘Five people/men.’ | b. tʃolo'vik pjetj
person(M).GEN.PL five.NOM?
‘Around five people/men.’ |
|---|--|

⁶² I argue for and against this extensively in (§5.3.).

Such a pattern is not only commonly shared by other Slavonic languages, but according to Greenberg's 44th Generalisation (1978: 284), it is a frequent cross-linguistic phenomenon.

The morphosyntactic agreement of these forms is a lot more straightforward than for prenominal numerals. The noun always takes GEN PL although there is more freedom for the verbal agreement.⁶³ If the verb precedes the numeral, it is likely to appear in SINGULAR (in the PRESENT tense); but if the verb appears after the numeral, most likely it will have PLURAL agreement. The same rule seems to apply to the PAST tense, with the exception that instead of PLURAL agreement, most of the time it will be NEUTER SINGULAR.

(16) (T1.1 24:03)

tam	bu-u	tʃolo'vik	moʒe	tʃitʃa
there	be-PST.M.SG	person.GEN.PL	maybe	thousand

‘There were perhaps a thousand people’.

Due to time limitations, I will focus mainly on numerals appearing before the NP in West Polesian, particularly lower and higher cardinals. Nevertheless, I give some details of postnominal numerals when there is a phenomenon potentially affecting both prenominals and postnominals (such as suppletion).

4.3. Collective numerals

As in other Slavonic languages, there are collective numerals in West Polesian. Nevertheless, their frequency in use and their morphosyntactic possibilities differ

⁶³ I have only found one exception to this in the entire corpus in which the speaker used a GEN SG.

considerably within the Eastern Slavonic sub-family, as I explain in this section. After setting a definition (§4.3.1.); I present the morphological and syntactic properties of collective numerals (§4.3.2.). After this, I explain the restrictions affecting the use of collective numerals happening at different levels (semantic, morphosemantic and morphosyntactic) (§4.3.3.). I finish this section with a summary and a comparison of collective numerals in Eastern Slavonic (§4.3.4.).

4.3.1. Definition

The main function of the collective numerals is to denote an (indivisible) group of items (most often animates), while emphasising the unity between them. Corbett (2000: 119) says that “[t]he primary function of collectives is to specify the cohesion of a group, sometimes manifested in joint activity”. See example (17).

(17) (Z.4.1.200:49)

z	nas	fest	duŕ:	bu-l-o	dit-ej	jestero ,
from	1SG.GEN	six	person.GEN.PL?	be-PST-N.SG	child-GEN.PL	six(COLL)
i	dvoje	batjk-o	z	matr-iju		
and	two(COLL)	father-NOM.SG	with	mother-INS.SG		

‘Altogether, we were six people: there were six children [together], plus (two) mother and father’.⁶⁴

Kim (2009) made a detailed analysis of collective numerals, where he compared collective numerals in BCMS, Polish and Russian from a Minimalist perspective

⁶⁴ I am aware that the maths do not match. However, for the sake of keeping the naturalness of the utterances as much as possible, I have not corrected those small calculation mistakes.

(particularly emphasising the Minimal Lexical Representation (MLR) and the Quantitative Case). I feel he gives a more complete definition than mine, as it is based on his cross-Slavonic survey and is comprehensive of the specificity of the different uses of collective numerals in Slavonic:

“They [Slavonic collective numerals] differ in that cardinal numerals are used for counting individual entities, whereas collective numerals are used for conveying a collective meaning, as well as some more idiosyncratic purposes, such as specifying mixed gender specification of a group, quantifying plural-only nouns, expressing stylistic differences, etc. [...]

The Russian collective numerals differ from the cardinal numerals in that the former emphasize ‘the totality’ or ‘the aggregate as a whole’, while the latter – ‘the individuated quantity’ (Bulaxovskij 1958; RG-I 1982; Suprun 1959; Timberlake 2004; Vinogradov 1947)” (Kim 2009: 1, 23) [See references there].

4.3.2. Morphological and syntactic properties

The form of the collective numerals is directly derived from cardinal numerals [*infra*] with the suffixes *–o(j)e* for numeral forms ending in a vowel (i.e. ‘two’ *dva* > *'dvoje*; ‘three’ *tri* > *'troje*). Stems ending in a consonant take *–ero/iro* (i.e. the rest of the numerals up to ‘ten’ or ‘twenty’, depending on the limit that we set) (e.g. ‘six’ *šest(j)* > *'šestero*).⁶⁵ According to Sussex & Cubberley (2006: 467) such suffixes are also found in all the Slavonic varieties except for Bulgarian and Macedonian. According

⁶⁵ Sussex & Cubberley (2006: 467) say that all the Slavonic 4+ numerals derive their collective form with the suffix *–(e)ro/(o)ro*, but I prefer to frame it in terms of morphophonological rules, rather than semantics.

to Suprun (1961), they were used in Old Church Slavonic (henceforth, OCS) as well, although less prominently than in Contemporary Russian (CSR).

Table 7 Collective numerals and their cardinal equivalents

Cardinal	Collective	Gloss	Cardinal	Collective	Gloss
<i>dva/dvi</i>	<i>'dvoje</i>	‘two’	<i>sjem/sim</i>	<i>'sjemero??</i>	‘seven’
<i>tri</i>	<i>'troje</i>	‘three’	<i>'visim/'vosim</i>	<i>'vos(j)mero</i>	‘eight’
<i>'tjtri</i>	<i>'tjetvero</i>	‘four’	<i>'devetj</i>	<i>'devetero?</i>	‘nine’
<i>pjetj</i>	<i>'pjatero</i>	‘five’	<i>'desitj</i>	<i>'desetero/ dr'sjatero?</i>	‘ten’
<i>fest(j)</i>	<i>'festero</i>	‘six’			

4.3.2.1. Agreement

Unlike lower cardinal numerals, collective numerals govern GEN PL in the rest of the numeral phrase when they appear in DIRECT CASES.

- (18) (Tor1.35.el 04:28)
- | | | | |
|--------------|--------------------|--------------|----------------|
| troje | 'xlopts-uv, | troje | sr'n-iv |
| three(COLL) | boy-GEN.PL | three(COLL) | son-GEN.PL |
- ‘Three boys, three sons.’

Yet, the noun inside the numeral phrase can appear with the same special GEN PL forms which a noun would take with higher cardinal numerals (19).⁶⁶

⁶⁶ See more on the potential GREATER ADNUMERATIVE in (§5.3.).

(19) a. (Z2.4a 01:07)

na ruk-ax osta-l-o-sja **troje** 'djitok
 in hand-LOC.PL leave.PRF-PST-N.SG-REFL three(COLL) child.GRADNM/GEN.PL?
 '[He] was left with three children (lit. in his arms).'

b. (B10.1 06:24)

ſest tſolovik vſrozl-ix, i v nas bu-l-o
 six.NOM person.GRADNM? adult-GEN.PL and in 1PL.GEN be-PST-N.SG

troje **dit-ej** vſe.
 three(COLL) child-GRADNM/GEN.PL? already
 'Six adults [lit. adult people], and we already had three children.'

4.3.2.2. Special uses

In contemporary West Polesian, collective numerals are used as a simpler alternative to the complex morphosyntactic system of cardinal numerals. Nevertheless, this phenomenon is explained in more detail in (§5.2.2.1.).

Table 8 Example of forms proposed by B2, when prompted with images

'two'	'three'	'four'	NOM PL	Gloss
(dva) lo'si	(tri)'losi /('troje) lo'siv	(tſtiri) 'losja	lo'sij	'elks'

The fact that none of the forms given for lower numerals are the same (see discussion on the illusion of discontinuity in (CRITERION 8 in (§5.3.3.)) shows a high level of variation (and perhaps instability) in this cell. For this reason, the forms used with collectives reduce part of that complexity.

In the PRESENT TENSE collective numerals trigger PLURAL agreement in the verb. Conversely, in the PAST TENSE they trigger NEUTER SINGULAR verbal agreement (like higher cardinal numerals) (as in (19) a-b.).

4.3.2.3. Dislocation

The most common word order for collective numerals is to appear before the noun (or NP) they quantify. Apparently, because of their straightforward agreement, they allow a dislocation of the heading numeral from the rest of the numeral phrase.

- (20) (B8.1 05:41)
- narodı-l-i-sja djit-i. oni pravda i ni 3ıv-utj,
 be_born.PRF-PST-PL-REFL child-NOM.PL 3PL.NOM truly and NEG live-3PL
- oni poumira-l-i v mene. umer-l-o tʃ... troje
 3PL.NOM die.PRF-PST-PL in 1SG.GEN die.PRF-PST-N.SG [four] three(COLL)
- de..., troje. ono odn-a dotʃ osta-l-a-sa.
 [child] three(COLL) just one-NOM.F daughter.NOM.SG remain-PST-F.SG-REFL
- dvoje umer-l-o menʃ-ıx de't-ej.
 two(COLL) die.PRF-PST-N.SG small-GEN.PL child-GEN.PL

‘Children were born. To be honest, they are no longer alive, they passed away [lit. ‘they died to me’]. Three [together], f... passed away, three [together]. Only one daughter is alive. Two small children [together] passed away.’

As you can see in (20), there is a verb (*umerlo* ‘passed away’) between the collective numeral (head) and the NP (*detej* ‘children’). Utterances like this seem very uncommon (at least, I have not been able to find more examples in the corpus), yet, at least, they seem possible. There are more examples of dislocations with other types of NP, especially with postnominal possessors which I mentioned in (§3.3.). So, such types of construction, which may sound ungrammatical to speakers of other Slavonic languages, might be a lot more common (and acceptable) in West Polesian than it appears, especially in oral informal speech.

I have also documented a sentence in a free text (21), in which the order of the constituents is not the one we would expect, from observing the behaviour of other collectives in Slavonic languages (determiner + collective numeral + noun).

- (21) (B20.16 01:27)
- | | | | | | |
|---------------|----------------------------|--------------|-----------|---------------|------------------------|
| fiet-ɪ | ¹xloptsɪ | dvoje | do | dom-u | [...]pɪbɪfɪ-l-ɪ |
| this-NOM.PL | boy.NOM.PL | two(COLL) | to | home-IGEN?.SG | run.PRF-PST-PL |
- ‘These two boys [together] fled home.’

So far, I have not reached an empirically solid explanation for the dislocation, but I consider it is important to document it. I hope that further analyses in the future will be able to account for it.

4.3.3. Restrictions

I deal with three types of restrictions affecting the use of collective numerals. From purely semantic restrictions (§4.3.3.1.) to morphosemantic restrictions (§4.3.3.2.), and morphosyntactic restrictions (§4.3.3.3.).

4.3.3.1. Semantic restrictions

4.3.3.1.1. Continuity

Since collective numerals denote a collective or group, at least in their primary sense, they cannot be employed to refer to two items happening separately (22).

(22) a. (B6&B9.el)

*vin	tak	zrobi-u	troje	ra'z-ov.
3SG.NOM.M	this way	do.PRF-PST.M.SG	three(COLL)	time-GEN.PL

(B6 & B9.vol) > > [tri ra'zi]
 three.NOM time.ADNM

‘He did this three times.’

b. *priʃli dvoje bra't-iv, odin vtʃora,
arrive.PRF-PST-PL two(COLL) brother-GEN.PL one.NOM.SG.M yesterday

a druŋi-ŋ tapera (B6 & B9.vol) > > [dv-a 'brati]
and second-NOM.SG.M today two-NOM.M brother.ADNM

‘Two brothers arrived; one yesterday and one today.’

Nevertheless, B6 and B9 approved (23), admitting that B6 proposed *tri ve'silje* 'three weddings' (with a cardinal numeral) as an alternative (and thus, probably more acceptable).

(23) (B6&B9.el)

v naʃ-emu sil-i bu-l-o **troje** **vese'l-ej**
in our-DAT.SG village-LOC.SG be-PST-N.SG three(COLL) wedding-GEN.PL

v aprelj-e
in April-LOC.SG

‘There were three weddings in April in our village.’

Looking at the corpus of other speakers from the same village (Bahdanaŭka), I have found an utterance (24) (from another even older and monolingual speaker) where the collective form would be violating the ‘continuity law’.

(24) (B8.4 03:51)

ja **dvoje** **dr't-ej** rodi-l-a sam-a [...]
1SG.NOM two(COLL) child-GEN.PL deliver-PST-F.SG REFL-NOM.SG.F

‘I gave birth to two children on my own [...].’

As far as it can be deduced from the story, B8 did not have twins. Nevertheless, in her mind, all her children form a family unit, which perhaps motivated her to use the collective form.

4.3.3.1.2. Prestige

Speakers disapprove the use of collective numerals when they are referring to titles, functions or professions considered to be of high respect, as in Russian (Mel'čuk 1985: 383). Again, the forms in (25) were proposed to B6, with B9 and B22 in the back contributing to the judgments:

(25) (B6, B9, B22.el)

Approved forms

- a. **soldat** **spas** **troje** **di't-ej**
 soldier.NOM.SG save.PST.M.SG three(COLL) child-GEN.PL
 'The soldier saved three children.'
- b. **dvoje** **sol'dat-ov** // **dvoje** **sol'dat**
 two(COLL) soldier-GEN.PL two(COLL) soldier.GEN.PL
 'Two soldiers.'

Rejected forms

- c. ***dvoje** **'batjušek** (B6.vol) > > [**dv-a** **'batjuški**].
 two(COLL) priest.GEN.PL two-NOM.M priest.ADNM
 'Two (Orthodox) priests.'
- d. ***dvoje** **pros'viter-ov** (B9.vol) > > [**dv-a** **pros'viteri**]
 two(COLL) pastor-GEN.PL two-NOM.M pastor.ADNM
 'Two pastors.'⁶⁷
- e. ***pjatero** **prisi'datel-ej** Kol'hoz-a (B6.vol) > > [**pjat** **prisi'datel-ej**]
 five(COLL) head-GEN.PL Kolkhoz-GEN.SG five head-GEN.PL
 'Five heads of the Kolkhoz [State-owned collective farm].'

4.3.3.2. Morphosemantic restrictions

The inventory of collective numerals is a finite set; however, it is unclear where the limit is. Different speakers have pointed to different quantities. Polish and BCMS allow collective numerals from 'two' to 'ninety-nine' (Kim 2009: 136), whereas Russian only allows 'two' to 'ten'. Some Ukrainian and Belarusian grammars (Pugh & Press 1999,

⁶⁷ NB. Notwithstanding the English cognate 'presbyter' seems a more accurate translation, it must be said that in rural Western Polesie Presbyterian church structure is rare or unknown. The term *presviter* (or *prosviter*) is often used interchangeably (in rural areas) for Protestant church ministers fulfilling the role of what technically is a 'pastor' or 'reverend'.

Rusanovskij et al. 1986, Vjarxoŭ 1961)⁶⁸ say that BLM and ULM have collective numerals from ‘two’ to ‘thirty’, which seems a very odd cutting point.⁶⁹

West Polesian collectives are probably only available for numerals from ‘two’ to ‘ten’ (like collectives in Russian). When I suggested *pjatnatsatero* B22 and B9 said that this sounded strange and B22 added “*boljŭ disjatŭ* [people] *nŭ buvaje!*” ‘There are never more than ten [people]!’. Having said this B19 (about fifteen years older than B9 and B22, but brought up in the same village, and also male) seemed happy with *dvatsat dvoje konej* ‘twenty-two horses (together)’, as long as it was not used in the ACCUSATIVE. I should stress that I only obtained forms like this by direct elicitation. Otherwise, I have never documented any speaker producing a collective numeral form higher than ‘ten’ on a free text or overheard conversation.

⁶⁸ Hurski (1972: 187) says that Russian collectives are most frequently used with numerals ‘two’ to ‘five’ and that they are limited (i.e. semantically restricted) up to ‘ten’, whilst in BLM and ULM collective numerals “overpass the first decene”. However he does not give more details.

⁶⁹ The way the system of numerals works in (at least most) Slavonic languages is the first decimals, then ‘ten’ to ‘twenty’ (which are a single phonological and orthographic word, and which govern the same type of agreements as other higher numerals, when cardinal), and from ‘twenty’ on the system replicates the system of the first decimals. There is also rich cross-linguistic evidence of languages having counting bases in ‘ten’ (e.g. Mandarin) or ‘twenty’ (e.g. Basque). In Comrie’s (2013) survey of numeral bases, out of 196 languages, only 5 followed a different base (besides 4 others that use extended body parts). Thus, such a claim is typologically quite suspicious.

4.3.3.3. Morphosyntactic restrictions

4.3.3.3.1. ANIMACY restrictions

The general underlying rule that can be deduced in the light of the data is that only ANIMATE nouns can be governed by a collective numeral (26).⁷⁰

(26) (B6.47.el)

Approved forms

- a. v lis-i bifaj-et⁷¹ **troje** 'losj-ej
in forest-LOC.SG run-3PL three(COLL) elk-GEN.PL

‘Three elks are running in the forest.’

- b. kupi-l-i **troje** 'kon-ej
buy-PST-PL three(COLL) horse-GEN.PL

‘[They] bought three horses.’

Rejected forms

- c. *fjestero traxto'r-iv (B6.vol) > > [fjest traxto'r-iv]
six(COLL) tractor-GEN.PL six tractor-GEN.PL

‘Six tractors.’

- d. *nazbira-l-i **troje** 'banok (B6.vol) > > [tri 'banki]
collect.PRFX-PST-PL three(COLL) jar.GEN.PL three jar.ADNM

‘[They] collected three jars.’

- e. ***troje** kvi'tok
three(COLL) flower.GEN.PL

‘Three flowers.’

⁷⁰ Nevertheless, we have just seen that in (23) B6 approved (but not preferred) **troje vese'lej** ‘three weddings’.

⁷¹ 3SG does not always have a final *-t(j)*, and its absence can sometimes be used as a clue to distinguish it from its PLURAL counterpart. In any case, since the verb appears before the NumP there also seems to be more freedom for verbal agreement.

- f. *troje 'pljafek
 three(COLL) bottle.GEN.PL
 ‘Three bottles.’
- g. *troje du'b-ov
 three(COLL) oak-GEN.PL
 ‘Three oak trees.’
- h. *desetero ja'jits
 ten(COLL) egg.GEN.PL
 ‘Ten eggs.’

Having made this first claim, it must be said that ANIMACY is complicated in West Polesian as I have already introduced in (§3.4.). The criteria most widely used for analysing ANIMACY in Slavonic languages has been the treatment of the cells of the ACCUSATIVE and the agreement of the NP (Eckhoff 2015, Huntley 1980). If a certain noun has an ACCUSATIVE PLURAL (and also SINGULAR, if MASCULINE) syncretic with the NOMINATIVE (henceforth ACC = NOM) it will be inanimate; whereas, if the ACCUSATIVE is syncretic with the GENITIVE (henceforth ACC = GEN), and so are all the targets, the noun will be ANIMATE.

Nevertheless, in most varieties of West Polesian the PLURAL sub-paradigm in NON-HUMAN ANIMATES presents a split. Instead of taking the ACC=GEN expected form, they take ACC=NOM (see (Huntley 1980, Roncero 2015) and references there), like in (27).

- (27) (B12.7 00:34)
 koro'vi pas-vi-v
 cow.ACC = NOM.PL graze-PST-M.SG
 ‘[He] used to graze cows.’

In (27) B12 does not use *ko'rtv* or *koro'vej* which would be the proper ACC PL = GEN PL form, but the ACC PL = NOM PL. According to the traditional criteria for analysis these nouns would be considered INANIMATE (at least in the PLURAL). Nevertheless, nouns denoting animals can be headed by collective numerals, which proves that they are ANIMATE. Thus, West Polesian distinguishes three levels (instead of two) in its animacy hierarchy: INANIMATE - ANIMATE (NON-HUMAN) - HUMAN, with some fluidity in the intermediate level. Now, when it comes to ACCUSATIVE contexts (i.e. situations in which the numeral phrase is the direct object) collective numerals outlaw the use of NON-HUMAN nouns (in most instances). That is to say, only nouns denoting humans can fulfil other roles than subject (NOMINATIVE).⁷² However, this rule does not apply unequivocally. As a piece of evidence, I present several examples from constructions that have been accepted and rejected by people from the same village. I am aware that I am dealing with rather marginal areas of the system and that it is not surprising to find inconsistencies and doubts from speakers (28).

(28) (B6.43.el)

Approved forms

- a. autobus zabira-je **pjatero** **dr't-ej** do nas.
 bus.NOM.SG collect-3PL five(COLL) child-GEN.PL to 1PL.GEN

‘The bus takes five children to our [school].’

- b. prisidatelj vibra-v **troje** 'xlopts-uv
 head.NOM.SG choose.PRF-PST.M.SG three(COLL) boy-GEN.PL

‘The head of the Kolkhoz chose three children.’

⁷² With the exception of pluralia tantum nouns, which can be used in ACC = NOM, probably because there are no other alternatives (given that they cannot be used with cardinal numerals).

But this does not explain why sentences like b., c. were accepted.

- | | | | | |
|----|---------------|------------------|--------------|------------------------------|
| na | Pask-u | zarīza-l-i | dvoje | kozle¹njat |
| in | Easter-ACC.SG | slaughter-PST-PL | two(COLL) | lamb.GEN.PL |

(B9.vol) > > [dv-i kozle¹njati]
 two-NOM.F? lamb.ADNM

b.	vona	pasvi-l-a	dvoje	'kon-ej ⁷³
	3SG.NOM.F	graze-PST-F.SG	two(COLL)	horse-GEN.PL

c.	kupi-l-i	troje	'kon-ej
	buy-PST-PL	three(COLL)	horse-GEN.PL

d. [B9 accepted it as a possibility; B21 rejected it from the beginning]

#Kolxoz	kupi-v	troje	tʃorn-ix	'kon-ej
K-NOM.SG	buy-PST-M.SG	three(COLL)	black-GEN.PL	horse-GEN.PL

e.	*na	dvori	pasvi-l-i-sj	dvatsat troje	'kon-ej
	in	outside	graze-PST-PL-REFL	twenty-three(COLL)	horse-GEN.PL

(B6.vol) > > [dvatsat tri ko'nɟɪ]
twenty-three horse.ADNM

‘There were twenty-three horses grazing outside.’

106

So far, the only sensible thing that can be concluded is a reiteration of what I have already said. There are some observable tendencies to restrict the use of collective numerals with non-human animates (particularly in the ACCUSATIVE). Yet, this being a marginal field of the language the norm has a less strict application than in other areas of the language, and thus the inter- and intra-speaker variability.

4.3.3.3.2. GENDER restrictions

In Contemporary Standard Russian (CSR) collective numerals are mostly restricted to VIRILE (i.e. HUMAN MASCULINE) nouns, nouns denoting young (MASCULINE) animals and *pluralia tantum* nouns (Kim 2009, Mel'čuk 1985, Pereltsvaig 2013).

CSR (In Pereltsvaig (2013: 310) modified)

(30)	a. ^I	dv-e	devušk-i
		two(CARDINAL)-NOM.F	girl-GEN.SG
	a. ^{II}	# dvoje	devušek
		two (COLL)	girl.GEN.PL
		'Two girls.'	

In the Russian National Corpus (2017), I have only found two hits for *dvoje devčat*, but no instances of **dvoje devušek*. In a similar vein, I have only found two hits for *dvoje ženščin* 'two women' in the corpus (in comparison to 717 hits for *dve ženščiny*, just in the NOMINATIVE).⁷⁴ Thus, we can deduce that such a construction is not

⁷⁴ Or even *dve sestry* 'two sisters' 314 hits (just in nom), but only 2 hits with the collective form *dvoje sester*.

completely forbidden, but certainly is disliked in CSR. Mel'čuk (1985: 382), points out that traditional grammars prescribe the use of collectives with FEMININE nouns, although the author himself recognises liking certain sentences containing a FEMININE with a collective (especially for nouns like *duša* 'person; soul'). However, he also admits that Reformatskij did not approve the same sentences. In short, there is no clear consensus about it in Russian but the most standard forms of the language restrict the use of FEMININE nouns with collective numerals.

West Polesian, as well as Standard Ukrainian, do not seem to be limited by this semantic constraint. Certainly, it is possible to find examples of collective numerals with MASCULINE and FEMININE nouns in West Polesian. For example, when I elicited the noun for 'eyebrow(s)' from B6 she gave [NOM SG] *brova* (F); [NOM PL] *'brovr*; but [2] *dvoje 'brovej*.

However, example (31) suggests that FEMININE nouns with collective numerals are not particularly well accepted by all.

- (31) (Tor 1.4 07:34)
- | | | | | |
|-----------------|-----------|------------------|----|-----------|
| dvoje... | dʒjv-i | 'dotʃkɪ | u | jɪ'ji |
| two(COLL) | two-NOM.F | daughter(F).ADNM | in | 3SG.GEN.F |
- 'Two [together]... she has two daughters.'

Hurski (1972: 185-187) remarks a specificity of BLM and ULM not present in CSR. Besides the fact that collective numerals are used a lot more frequently in BLM and ULM, they use collective numerals (in free variation with cardinals) with nouns containing the suffix *-in/yn* (e.g. *balharyn* 'Bulgarian') and young creatures (ANIMATES).

According to Kim (2009: 135), BCMS, Polish and Russian collective numerals can be used to quantify or specify a mixed-gender group and pluralia tantum nouns. The same rule applies to West Polesian; e.g. (B6) *dvoe nafia'vits* ‘two trousers’. Kim (ibid) also points out that collective numerals can be used to quantify infants or young animals in BCMS, Polish and Russian. According to Hurski (1972), this also happens in BLM and ULM, so West Polesian does not seem an exception in the Eastern Slavonic sub-family (e.g. *troe dıtaj/ditej* ‘three children’). However, this may be due to the fact that West Polesian collective numerals allow you to quantify any animate noun with a collective. In other words, the fact that such class of nouns can be used with a collective numeral is rather a side effect of the extension of uses of collective numerals to all ANIMATES. Mel’čuk (1985) also includes deadjectival nouns (in Russian) as a noun class which can only be quantified with a collective numeral. I am not sure whether there is such a restriction in West Polesian, but it is certain that collective numerals are used with deadjectival nouns (32):

- (32) (Xab1.6 00:41)
 id-utj **dvoje** **mal-ix**
 go-3PL two(COLL) young-GEN.PL
 ‘Two youngsters (lit. youngs) go.’

4.3.3.3.3. CASE restrictions

As far as data from the corpus reveal, collective numerals are only available for direct cases (i.e. NOM-ACC). This should not be surprising, as it is a tendency in most (if not all) Slavonic languages. Kim (2009) states that BCMS allows inflecting collective numerals in other cases, but in reality, this is hardly done, especially in spoken language. The situation for Polish is the same as in West Polesian:

“From these old forms of the agreeing collective numerals, only neuter singular forms survive in contemporary Polish. They are found only in the nominative and accusative cases [...]” (Kim 2009: 138).

On the one hand, Mel’čuk (1985: 439) gives examples of the ACCUSATIVE form of the collective numerals (vs. the usual NOM form), although he does not give any examples of collectives being used in non-direct cases (as he focuses primarily on cardinal numerals) (33). This should make us suspect that collective numerals in NON-DIRECT CASES are also outlawed in Russian.

(33) **CSR** (In Mel’čuk (1985: 439))

Vrač	osmotre-l	tro-ix	bol’n-yx
doctor.NOM.SG	examine.PRF-PST.M.SG	three(COLL)-ACC.PL	sick-ACC.PL

[* tro-e	bol’n-yx]
three(COLL)-NOM.SG	sick-ACC.PL

‘The doctor examined three patients (lit. sick).’

On the other hand, Vjarxoŭ (1961: 24) gives the full paradigm of the collective numerals ‘two’, ‘three’ and ‘four’ in BLM, CSR and ULM. He admits that (at least for BLM) they are very rarely used in oblique cases, and he does not even provide any example of their use in Contemporary BLM.

Based on observations from the corpus of West Polesian, when a non-direct case is required, the forms of the cardinal numerals are employed instead. This also seems to be the common practice in the other three Eastern Slavonic varieties, even when a potential dedicated collective form may exist (perhaps more as ‘possible words’).

4.3.4. Summary

I have summarised the properties and behaviour of collective numerals in Eastern Slavonic and West Polesian in Table 9:

Table 9 Collective numerals across Eastern Slavonic

	Belarusian	Russian	Ukrainian	West Polesian
No GENDER restrictions	+	–	+	+
No semantic restrictions:				
· HUMAN nouns	+	+	+	+
	(especially if in <i>–in</i>)		(especially if in <i>–in</i>)	
· Young animals	+	+	+	+
· ANIMATES	–	–	?	+
· Concrete INANIMATES	–	–	–	?
	(only PL tantum)	(PL tantum, optionally)	(only PL tantum)	(probably, only PL tantum)
· Abstract nouns	–	–	–	dispreferred
· Prestigious titles	–	–	?	–
Morphosemantic restrictions	Only from ‘two’ to ‘thirty’	Only from ‘two’ to ‘ten’	Only from ‘two’ to ‘thirty’	Probably, only from ‘two’ to ‘ten’
Morphosyntactic restrictions	Virtually limited to DIRECT CASES	Virtually limited to DIRECT CASES	Virtually limited to DIRECT CASES	Only available for DIRECT CASES
Subject- Predicate Agreement	NEUTER SINGULAR/ PLURAL	NEUTER SINGULAR/ PLURAL ⁷⁵	NEUTER SINGULAR/ PLURAL?	NEUTER SINGULAR/PLURAL?

⁷⁵ (Kim 2009: 188).

In this section, firstly, I define pronominal numerals and I show how they differ from other types of numerals in West Polesian (§4.4.1.); secondly, I provide an analysis of the properties of this particular class of numerals (§4.4.2.), especially focusing on its ANIMACY constraints; and thirdly, I compare pronominal numerals with similar numeral classes (and subclasses) in other Slavonic languages (§4.4.3.). My conclusion is that West Polesian pronominal numerals are very special cross-linguistically (§4.3.4.).

4.4.1. Definition

Pronominal numerals are a (minor) class of numerals with a double function: anaphoric and quantificational. That is to say, they replace a noun or an NP whilst also specifying the number of entities (humans) involved.

Compare also sentences (35) a., a proper pronominal numeral, with (35) b., a collective. They were all produced by the same speaker within the same free text.

(35) a. (B6.11 05:12)

jix	tʃetr'rojx	bu-l-o....,	de't-ej
3PL.GEN	four(PRON)	be-PST.N-SG	child-GEN.PL

‘There were four [of them]..., children.’

b. (B6.11 03:05)

nas	'tʃetvero	dɪ't-ej	bu-l-o
1PL.GEN	four(COLL)	child-GEN.PL	be-PST.N-SG

‘We used to be four children [in total].’

Even though there may be some common origin, pronominal numerals are different from the ACC = GEN form of the cardinal numerals.

- (36) a. *kupi-v dvox ko'rovi (B9.vol) > > [dʒvi ko'rovi]
buy-PST-M.SG two(PRON) cow.ADNM two.NOM.F cow.ADNM
'[I] bought two cows.'
- b. *kupi-v dvox buha'j-ru (B6&B9.vol) > > [dva buha'ji]
buy-PST-M.SG two(PRON) bull-GEN.PL two.NOM.M bull.ADNM
'[They] bought two bulls.'

Looking at the examples in (34) and (36) we can see they are different from *dvoje*, which is a collective numeral; as well as different from what is often called ‘distributive’ *oba/obidva/oboje* ‘both’ (§4.5.3.).

Having said this, I found a speaker making an intriguing use of pronominal numerals. She seems to distinguish the collective numeral (^l**trojr**) from the ‘regular’ pronominal numerals (**dvox**). However, it is unclear what the forms like **dvo^ljix**, **tro^ljix** (note also the stress shift) are, and to which I have not attested elsewhere. Syntactically they behave like pronominal numerals, but the form **dvox**, which exists in her inventory, is already a pronominal (37).

- (37) a. (T7.5 00:11)
- | | | | | |
|-----------|-------------|---------------|---------------|----------------|
| ʃ:je | z | jim | robi-l-o | dvoʃix, |
| even | with | 3SG.INS.M | work-PST-N.SG | two(PRON) |
| | | | | |
| dv-a | 'xloptsa | robi-l-o | | |
| two-NOM.M | boy.GEN.SG? | work-PST-N.SG | | |
- ‘And two worked with him, two boys worked with him.’

b. (T7.6 00:56)

mi z Raj-u **dvox** misi-l-i
1NOM.PL with Raja-INS.SG two (PRON?) dare? -PST-PL

‘Me and Raja, we both (lit. the two) dared [obscure meaning].’

c. (T7.15 00:03)

a tam bu-l-o **tro'jix**, 'troji dɪ't-ej ʊ jɪɪ
and there be-PST-N.SG three(PRON) three(COLL) child-GEN in 3SG.GEN.F

sjim-i bu-l-o
family-GEN.SG be-PST-N.SG

‘And there were three, her family had three children.’

There are not enough tokens to make any significant conclusions, so I can only propose two possible solutions. The first possibility would be that both *dvox* and *dvo'jix* are variants of the pronominal numerals, *dvo'jix* probably being a more locally restricted variant. The second possibility would be that forms like *dvo'jix* are a subclass of pronominal numerals. Given that the idea of a sub-class of pronominals (which are a very small class *per se*) seems less likely, I favour the first possibility. In any case, data from the corpus prove that pronominal numerals are a morphologically dedicated (and distinct) class in West Polesian.

4.4.1.1. Geographic specificity

I have documented instances of what looks like pronominal numerals in most of the villages I have been working at. Nevertheless, I have not been able to find many examples in the corpus of recorded texts from other varieties (villages), but I can say

that pronominal numerals are a morphologically robust category, at least in the variety of Bahdanaŭka (Luninec). I only realised that pronominal numerals were different from the rest of the numerals (particularly COLLECTIVES) by the end of my second expedition, which was solely conducted in Bahdanaŭka, with B6 as my main language assistant. For this reason, there was little time for elicitation sessions on this parameter and so most data come from free texts.

4.4.2. Properties

Hereafter, I describe the properties of pronominal numerals by explaining the different restrictions they present at different levels. I start with morphosemantic restrictions (§4.4.2.1.). Then I move to purely syntactic restrictions (§4.4.2.2.) and morphosyntactic restrictions (§4.4.2.3.), related to GENDER (§4.4.2.3.1.); NUMBER (§4.4.2.3.2.); ANIMACY (§4.4.2.3.3.); and CASE (§4.4.2.3.4.). I end up with restrictions affecting the presence of adjectives (§4.4.2.4.).

4.4.2.1. Morphosemantic restrictions

Pronominal numerals seem to be morphosemantically restricted to numerals from ‘two’ to ‘ten’. Nonetheless, in reality forms like *dvox* and *trox* are a lot more common than the rest. For that reason, I have not been able to document all the existing forms, yet I have a strong intuition that there is a form for the grids that are still empty (Table 10).

Table 10 Pronominal numerals

2	3	4	5	6	7	8	9	10
dvox	trox	tʃtʲr'ox/ tʃetʲr'jox	pja'tjox	ʃis'tjox	n.d.	n.d.	n.d.	dʲisi'tjox

Having said this, B9 said that *pjatnatsaterox* ‘the fifteen’ is possible, but both B9 and B22 agreed it was very weird (“*nʲi buvaje stiljke*” ‘There are [never] that many’).⁷⁷ Thus the semantic restriction is not entirely strict, although it seems related to a more general underlying phenomenon in the language (and also well-attested in Russian): INDIVIDUATION. In other words, the higher the amount, the harder it is to treat the entities individually. Data from West Polesian point out that pronominal numerals are [+SPECIFIED], especially in comparison to the collective numerals (where there is less specification). So if SPECIFICATION is heavily dependent on individuation, it should be not surprising that the speakers disprefer the use of pronominal numerals with quantities higher than ‘ten’. Nevertheless, it is important to remark that there is no morphologically dedicated form for ‘one’. The cardinal form (which is quasi-adjectival, (§4.2.1.1.)) can be used on its own as a pronoun, but it cannot be considered a pronominal numeral *stricto sensu*.

4.4.2.2. Syntactic restrictions

Due to their pronominal nature, they do not allow the presence of their antecedent in the sentence, as in (34) c. Apart from (35) a., where there is a pause (creating a type

⁷⁷ B6.28 01:20

of afterthought); so far, I have only found an exception in the corpus where a pronominal numeral appears with its antecedent (38).

- (38) (B20.15 00:01)
- | | | | | | |
|--------------------------|--------------------------------|----|------------------------------|-------------------|------------|
| kosi-l-i | dvox ⁱ, | to | jestj, | nakosi-l-i | sin-a |
| reap-PST-PL | two(PRON) | so | be.PRES | reap(PRF?)-PST-PL | hay-ACC.SG |
| dvox ⁱ | molod-ije | | 'xloptsr ⁱ | | |
| two(PRON) | young-NOM.PL | | boy.NOM.PL | | |

‘Two were reaping..., that is to say, two young boys were reaping hay.’

In (38) it seems that B20 was aware of the specificity of *dvox* and suddenly realised that I may not be able to understand it. For that reason, he gave an explanation, which was not be very grammatically correct, but which underlined what he was referring to by using this pronominal form. Otherwise, the rest of the data in the corpus adjust to this syntactic restriction.

4.4.2.3. Morphosyntactic restrictions

4.4.2.3.1. GENDER restrictions

They are available for nouns of any GENDER. But because pronominal numerals are conditioned by ANIMACY, as I explain in (§4.4.2.3.3.) the only NEUTER noun that fulfils the criteria is *ditjo/djtja/ditna* ‘child’. Otherwise, the rest of the nouns replaceable by a pronominal numeral are either MASCULINE or FEMININE (and I have documented enough examples of FEMININE and MASCULINE nouns being replaced by a pronominal numeral). The sentences in (39) were accepted with no objection by B5, B6, B9 and B22.

- (39) a. [tʃɪtelʲk-ɪ] trox vʒe na pensij-u,
 teacher-NOM.PL three(PRON) already in pension-LOC.SG
 a ostaljn-ija ʃtʃe rob-ljat.
 and rest-NOM.PL still work-3PL
 ‘[The female teachers] three of them are already retired, but the rest are still working.’
- b. [dotʃkɪ] odn-a tut, a dvox v Pinsk-e rob- ljatj
 daughter.NOM.PL one-NOM.F here and two(PRON) in Pinsk-LOC.SG work-3PL
 ‘[X’s daughters] one works here, and the [other] two in Pinsk.’

4.4.2.3.2. NUMBER restrictions

Pluralia tantum nouns cannot be replaced by pronominal numerals. The reason behind this seems not that much related to NUMBER as such, but mostly (or solely) motivated by the strong ANIMACY restrictions for pronominal numerals (§4.4.2.3.3.). Having said this, Greville G. Corbett (p.c.) and I tried looking for instances of pluralia tantum nouns that refer to humans in European languages, but we could only find a couple (and none of them in Eastern Slavonic). Thus, since pluralia tantum nouns are almost inherently INANIMATE, it is unlikely that any pluralia tantum noun can be substituted by a pronominal numeral. The speakers B6 and B9 rejected the forms in (40), and in some cases, they offered the noun with a collective numeral instead.

- (40) a. *dvox tʃaʲsɪ ‘two hours/ clocks’⁷⁸ b. *dvox panʲtʃoxɪ ‘two pairs of socks’
 c. *dvox nahaʲvɪts ‘two pairs of trousers’

⁷⁸ I realised once I was back from the field that in the variety of Bahdanaŭka the preferred word for ‘watch, clock’ is *zefiarkɪ* (probably a loanword from Polish *zegarek*). That could have caused them to dislike *tʃasɪ*. Yet, it is clear from other examples that the form *zefiarkɪ* would have been equally discarded due to ANIMACY restrictions (§4.4.2.3.3.).

4.4.2.3.3. ANIMACY restrictions

Only HUMAN ANIMATE nouns can be replaced by a pronominal numeral. Examples

(41) a-c. were rejected by B6.

- (41) a. A: – **tabu¹retk-i** **jestja?**
 stool-NOM.PL be.PRS

B: – * **tak, ma-em trox.**
 yes have-1PL three(PRON)

A: – Have you got [any] stools?

B: – Yes, we have three.

- b. A: – **maʃin-i** **jestja?**
 car-NOM.PL be.PRS

B: – * **da, dvox.**
 yes two(PRON)

A: – Have you got any cars?

B: – Yes, (we have) two.

- c. A: – **kɨljkɨ dnj-ej lɨʒa-l-a v bolnits-e?**
 how_many day-GEN.PL lie-PST-F.SG in hospital-LOC.SG

B: – * **dvox.**
 two(PRON)

A: – How many days did [she] stay at the hospital?

B: – Two.

NON-HUMAN ANIMATES (usually farm animals) cannot be replaced by a pronominal numeral. B6 and B9 rejected sentences like the ones in (42) and they volunteered different forms instead, some of which contained collective numerals. In theory, the use of collective numerals is outlawed when they are determining a NON-HUMAN

ANIMATE noun which stands in ACCUSATIVE. So far, my explanation is that as long as the collective numeral stands on its own it is acceptable, but not preferred.⁷⁹

(42) (B6&B9.el)

- a. *[tse¹zark-i] u nas **dvox.**
 guineafowl-NOM.PL in 1PL.GEN two(PRON)

‘[Guineafowls]⁸⁰ we have two (of them).’

- b. *¹svinj-e / vipruv¹k-i derz-imo; utfra bi-l-i **dvox.**
 pig-ACC.PL boar-ACC.PL farm-1PL yesterday hit-PST-PL two(PRON)

(B9.vol) > > [utfra bi-l-i **dvoje]**
 yesterday hit-PST-PL two(COLL)

‘We farm pigs/boars; yesterday we slaughtered two (of them).’

4.4.2.3.4. CASE restrictions

Data from fieldwork suggests that pronominal numerals have a defective CASE infection, although it is clearer for some CASE values than for others. As a result, in OBLIQUE CASES (at least, in certain) pronominal numerals are replaced by the cardinal numerals. There seems to be enough evidence for stating that, besides the NOMINATIVE, pronominal numerals can be used with the VOCATIVE (43) a., and the ACCUSATIVE (43) b.⁸¹ However, with the DATIVE (43) c, they seem to be defective and use the CARDINAL form (in the DATIVE) instead:

⁷⁹ Read more details on collectives and animacy restrictions in (§4.3.3.3.1.).

⁸⁰ *Tsezarka* is a type of domestic bird, more specifically, a ‘helmeted guineafowl’.

⁸¹ I admit I lack data reliable data for the GENITIVE forms.

(43) (B6,B9.el)

- a. ej, vi **dvox!**
 hey 2PL.NOM/VOC two(PRON).VOC
 ‘Hey, you two!’
- b. ni batʃ-u **fiet-ix dvox**
 NEG see-1SG this-ACC.PL two(PRON).ACC
 ‘I can’t see these two.’
- c. A: kɪljɪk-ɪm xlopts-am dava-l-i mɪdalj?
 how_many-DAT boy-DAT.PL give-PST-PL medal.ACC.SG
- B: **dvo-m, dvo-m xlopts-am**⁸²
 two(CARD)-DAT two(CARD)-DAT boy-DAT.PL
- A: How many boys/kids did they give the medals to?
 B: To two; to two kids.

When it comes to the LOCATIVE and INSTRUMENTAL cases, the system itself presents some problems. In the variety of Bahdanaŭka, where I discovered and conducted all the specific fieldwork on pronominal numerals, the LOCATIVE case cannot be used with any HUMAN noun. All the prepositions that govern LOCATIVE that go with a HUMAN NP in other neighbouring Slavonic varieties (including other varieties of West Polesian) (i.e. [BLM] *u*; [CSR] *o*; [ULM] *pry*) are not permitted in this variety. Since the LOCATIVE is a syntactically non-autonomous CASE value,⁸³ it cannot be used unless there is a preposition (that is why in the Russian descriptive tradition it is often referred to as PREPOSITIONAL). When I elicited examples like (44) I was not fully aware of the

⁸² We have seen that the pronominal numerals do not allow the presence of their referent in the same phrase, for this reason, this form can only be a cardinal numeral.

⁸³ Using the terminology in (Brown 2007, Corbett 2012).

ANIMACY restrictions for the pronominal numerals, but I knew that no HUMAN noun could ever appear in LOCATIVE, that is why I suggested that sentence. Not surprisingly, the speakers flagged it as ungrammatical:

(44) (B6.el)

A: v kiljk-ix xat-ax osta-l-i-sja viteran-i?
in how_many-LOC house-LOC.PL remain-PST-PL veteran-NOM.PL

#B: vo dvox
in two(PRON).LOC?

(B6 vol) > > [vo dvox xat-ax]
in two(PRON?).LOC house-LOC.PL

A: In how many houses are they still veterans?

B: In two; in two houses.

The INSTRUMENTAL also presents challenges. In all the corpus I have not been able to identify any instances of it, yet I have a strong intuition that forms like **dvoxju* or **dvoxom* are ungrammatical. Conversely, I have identified the form *dvox* (which in other syntactic contexts is the pronominal numeral) being employed as an INSTRUMENTAL, not only in the variety of Bahdanaŭka (45) a., but also in Žydča/Siamikhavičy (45) b-c.

(45) a. (B6.11 02:40)

mi dvox s sistr-oju spa-l-i
1PL.NOM two(PRON) with sister-INS.SG sleep-PST-PL
‘The two of us, me and my sister used to sleep [together]...’

b. (Z1.2 0:37)

staralısjə,	robi-l-i	s	sis'tr-oju	dvox
try-PST-PL-REFL	work-PST-PL	with	sister-INS.SG	two(PRON?)

‘We did our best, I used to work [hard] with my sister.’

c. (Z4.5 08:18) [Z1 SPEAKING]

to ja	z	Nadij-u	dvox	bu-l-a	i	mat-i
so 1SG.NOM	with	Nadja-INS	two(PRON)	be-PST-F.SG	and	mother-NOM.SG

‘We were two with Nadyja, and then mum.’

Thus, it is most likely that pronominal numerals are defective, although they can fulfil at least certain syntactic roles in the DIRECT CASES.

4.4.2.4. Restrictions with (attributive) adjectives

The speakers B6 and B9 accepted the examples in (46) and, (47) admitting that not all sound completely natural:

(46) a. (B6.el)

u	Manj-i	pjatj	dɪ't-ej;	star-ije	u	fiorad-ax,
in	M.-GEN	five	child-GEN.PL	old-NOM.PL	in	city-LOC.PL

a	trox	menʃ-ix	tutaj.
and	three(PRON)	young-GEN.PL	here

‘Manja has five children. The oldest ones are in the city, and the three youngest are here.’

b. (B6, B9.el) [Situation: People from Ukraine have moved to our village].

dvox **nov-ix** tut ʒiv-utj; a ostaljn-ija tam;
two(PRON) new-GEN.PL here live-3PL and rest-NOM.PL there

B6.vol > > [**dvox** tut ʒiv-utj]
two(PRON) here live-3PL

‘Two of the new ones live here, and the rest there.’

Nevertheless, the following sentence was rejected by B6 and B9, for reasons still unknown to me:

(47) (B6.el) [Situation: I visited my neighbour last night. He has four children, but only the two youngest greeted me.]

***dvox** **mal-ix** pozdarova-l-i-sja; finʃ-ija ne
two.(PRON) young-GEN.PL greet-PST-PL-REF other-NOM.PL no

‘The two youngest greeted [me]; but the rest didn’t.’

The only exception to this in the corpus has been already mentioned (38), and as I said, it seems that B20 was trying to explain the meaning of *dvox* to me. In fact, other people from the same village (B9, B21) rejected that sentence. Yet, they accepted (*zabrali dvox molodix xloptsuv* ‘they took two young boys’ because they are the direct object (ACCUSATIVE)).

So regardless of the fact that pronominal numerals with attributive adjectives may be ungrammatical or not, it is a fact that they are certainly disliked. Moreover, I have not been able to find any clear or convincing instance of pronominals with adjectives in predicative function in the corpus.

4.4.3. Cross-linguistic comparison

4.4.3.1. Pronominal numerals and other Slavonic languages

The etymology of pronominal numerals is still uncertain to me;⁸⁴ thus, I am not going to speculate too much about it. Yet, data from Common Eastern Slavonic and neighbouring Slavonic varieties point out that this is an innovation of West Polesian. To my knowledge, no data has ever been published on this phenomenon about any other Slavonic language. This does not exclude the possibility that they may exist in other Slavonic varieties, but most likely pronominal numerals do not constitute such a morphologically robust category as in West Polesian. In order to prove the specificity of this class of numerals, I will contrast different (better-attested) phenomena in Slavonic languages with data from West Polesian.

4.4.3.1.1. Quasi adverbial numerals

Certain Slavonic languages, such as Russian (CSR) and Belarusian (BLM), have a quasi-adverbial form of the numerals [CSR] *dvojom* or [BLM] *ŭdvaix* ‘the two of [us] (together)’. I deal specifically with these quasi-adverbials in (§4.5.4.). Nevertheless, in short, quasi-adverbial numerals seem morphosemantically restricted to numerals up to ‘ten’ with higher numerals (i.e. ‘five’ to ‘ten’) being extremely rare, even if they could exist in any person’s idiolect. Note that this construction is different from the

⁸⁴ Matthew Baerman (p.c.) suspects it is derived from the GENITIVE form of the cardinal, like Polish virile numerals. In West Polesian the GENITIVE form of the cardinal ends in *-ox* (many speakers use *-ux* for ‘two’); e.g. [NOM M] *dva*; [GEN SG] *dvox/dvux* ‘two’. According to Miechowicz-Mathiasen (2011: 2-3), the GENITIVE of Polish LOWER CARDINAL numerals is systematically syncretic with NOMINATIVE of VIRILE (LOWER) numerals; e.g. [NON-VIRILE NOM] *trzy*; [VIRILE NOM] *trzech*; [GEN (ALL GENDERS)] *trzech* ‘three’.

INSTRUMENTAL form of cardinal numerals. Compare (48) a.^I-a.^{II} (a cardinal in INSTRUMENTAL) with (48) b.^I-b.^{II} (quasi-adverbial).⁸⁵

(48) **BLM**

a.^I **z** **dzvu-mja** **sjabroŭk-ami**
with two-INS friend-INS.PL

CSR

a.^{II} **s** **dvu-mja** **podrug-ami**
with two-INS friend-INS.PL

‘With two [female] friends’

BLM

b.^I My **ŭdvaix**, z **sjabroŭk-aj**, **prynjas-l-i** **cukerk-i** **Maš-e**
1PL.NOM two(QADV) with friend-INS.SG bring.PRF-PST-PL sweet-ACC.PL M.-DAT.SG

CSR

b.^{II} My **dvojom**, s **podrug-oj** **prines-l-i** **konfetc-i** **Maš-e**
1PL.NOM two(QADV) with friend-INS.SG bring.PRF-PST-PL sweet-ACC.PL M.-DAT.SG

‘The two of us, with my [female] friend, brought sweets to Masha.’

This type of quasi-adverbial numerals also exists in West Polesian, and although they seem very closely related to proper pronominal numerals (because of the morphology), they are of a different nature:⁸⁶

⁸⁵ Many thanks to Anastasija Asjuk (p.c.) for checking this.

⁸⁶ Note that the same speaker (B20) has been documented using pronominal numerals (38).

(49) (B20.17 01:02)

i lud-i boja-l-i-sa tʃerez nifio i-ti,
and person-NOM.PL scare-PST-PL-REFL through 3SG.GEN.M go-INF

obizatelno ʃjob udvox tʃi utrox
necessarily COMPL two(QADV) or three(QADV)

‘People were afraid of crossing it [a haunted bridge], it was necessary [to do it] together with two or three people.’

Yet, quasi-adverbial numerals cannot appear as the subject of a sentence on their own.

That is to say, they require a pronoun in NOM, even if this one is PRO-DROP. Note that in the sentences (50) a.^I-a.^{III}, the subject is the 3PL NOM pronoun, not the numeral. The West Polesian equivalent of (50) a.^I, a.^{II} is (50) a.^{III}. Notice how (50) a.^{III} (QUASI-ADVERBIAL) contrasts with (50) b. (PRONOMINAL), an utterance accidentally volunteered by B13 in an elicitation session.

(50) **BLM**

a.^I Jany pajš-l-i v armij-u ŭdvaix
3PL.NOM go.PRF-PST-PL to army-ACC.SG two(QADV)

CRS

a.^{II} Oni poš-l-i v armij-u dvojom
3PL.NOM go.PRF-PST-PL to army-ACC.SG two(QADV)

WP

a.^{III} voni poʃ-l-i v armij-u udvox
3PL.NOM go.PRF-PST-PL to army-ACC.SG two(QADV)

‘The two together went to the army.’

b. (B13.vol)

dvox paʃ-l-i v armij-u
two(PRON) go.PRF-PST-PL to army-ACC.SG

‘(Some/The) two went to the army.’

4.4.3.1.2. Animate numerals

Some Slavonic languages, such as Polish (Brown 1998, Miechowicz-Mathiasen 2011), present a special form for the numeral when it is heading a noun phrase containing a HUMAN ANIMATE noun (i.e. acting as a controller).⁸⁷ See the examples (51) a-b. from Polish. Note that instead of the regular MASCULINE form of the numeral *dwa* ‘two’, Polish uses a special form *dwaj*. Failure to use the ANIMATE numeral in these contexts (particularly (51) a.) makes the sentences ungrammatical. We see that in (51) b. the numeral appears with a pronoun as well, but this has a restricted use. Both examples are based on *Harry Potter and the Chamber of Secrets* (J.K. Rowling, 1998) in (Waldenfels & Meyer 2011), and the Polish translation available there.

Polish/English

- (51) a. [...] Fred i George, **dwaj** **brac-ia** bliźniac-y Ron-a.
F.-NOM and G.-NOM two(ANIMATE) brother-GEN.SG twin-NOM.PL R.-GEN
“ [...] Fred and George, Ron’s elder twin brothers.”

- b. [...] czek-a, żeby usłysze-ć, dlaczego **wy** **dwaj**
wait-3SG COMP listen.PRF-INF why 2PL.NOM two(ANIMATE)
nie przyjecha-l-iście pociąg-iem razem z inn-ymi.
NEG arrive.PRF-PST-2PL train-INS.SG together with other-INS.PL

“ [...] he’s waiting to hear why you two didn’t arrive on the school train.”

Bear in mind that the main rule for using pronominal numerals in West Polesian is that the noun (or NP) being replaced cannot be present in the sentence. This means

⁸⁷ Many thanks to Mateusz Tworzewski (p.c.) for pointing this to me.

that (52) b.^I is not acceptable in Polish (in spite of (51) b. with *dwaj*),⁸⁸ and neither is (52) a.^{II} considered grammatical in West Polesian.⁸⁹

West Polesian

- (52) a.^I pja'tjox kupi-l-i xat-u sobi
 five(PRON) buy-PST-PL house-ACC.SG REFL.DAT
- a.^{II} *pja'tjox bra't-iv kupi-l-i xat-u sobi
 five(PRON) brother-GEN.PL buy-PST-PL house-ACC.SG REFL.DAT
- ‘(The) five (brothers) bought a house for themselves.’

Polish

- b.^I # Pięciu kupi-ł-o sobie chat-ę.
 five(ANIMATE) buy-PST-N.3SG REFL.DAT house-ACC.SG
- b.^{II} Pięciu brac-i kupi-ł-o sobie chat-ę.⁹⁰
 five(ANIMATE) brother-NOM.PL buy-PST-N.3SG REFL.DAT house-ACC.SG
- ‘(The) five (brothers) bought a house for themselves.’

Thus, whilst animate numerals belong to a subclass of cardinal numerals in Polish, West Polesian pronominal numerals comprise a morphologically separate class of numerals.

⁸⁸ N.B. The numeral does not appear with a MASCULINE noun, but with a pronoun.

⁸⁹ Many thanks to Paula Borowska (p.c.) for her contribution to this point.

⁹⁰ The expression *w pięciu* seems more acceptable (or at least, it provides some results in Google searches), but that is more a distributive circumlocution (i.e. they bought the house ‘between the five of them’) and it is not the pure syntactic subject of the sentence (and thus it is not in NOMINATIVE). Otherwise, Google searches give almost no results of *pięciu* used on its own.

4.4.3.1.3. Collective numerals as pronouns

In Contemporary Standard Russian (CSR) it is possible to use a collective numeral in contexts like the one in (53) where it behaves like a pronoun in syntactic terms.

(53) CSR (Russian National Corpus, 2017)⁹¹:

I totčas meždu rjad-ami zasnova-l-i **dvoe** — **devuš-k-a**
and then between queue-INS.PL move.IPFV-PST-PL two(COLL) girl-NOM.SG

i **junoš-a** oba s karandaš-ami i bloknot-ami, [...]
and youth-NOM.SG both with pencil-INS.SG and notebook-INS.SG

And in that moment between the queues two [people] were going back and forward: a girl and a young man, both with pencils and notebooks, [...].’

West Polesian also has collective numerals (§4.3.), and they can also be used in the same type of contexts (54) a-b. However, note that in (54) b. there is a clear contrast between the collective (**dvoje**, **troje**) and the pronominal numerals (**trox**).

(54) a. (P2.1 03:00)

u mana **dvoje**; **dotjka** i **sin**
in 1SG.GEN two(COLL) daughter-NOM.SG and son.NOM.SG

‘I have two: a daughter and a son’.

⁹¹ Rubina, Dina (2008-2009). *Belaja golubka Kordovy*.

- b. (P1.2.2.pr 00:01) [P1 was asked how many children were in her family]
- troje** **mi**; dv-a 'sina, I bu-l-o u jiji, ja,
 three(COLL)1PL.NOM two-NOM.M son.GEN.SG and be-PST-N.SG in 3SG.GEN.F 1SG.NOM
 dotʃk-a u mam-I bu-l-a; **trox** ano mi. I
 daughter-NOM.SG in mum-GEN.SG be-PST-F.SG three(PRON) but 1PL-NOM and
 u mene **troje** bu-l-o; odn-ofio ni-ma, a **dvoje** ʃtʃe.
 in 1SG.GEN three(COLL) be-PST-N.SG one-GEN.SG.M NEG-HAVE and two(COLL) still
- ‘Altogether we [were] three: two sons, and mum also had a daughter; just us three. And I also had three; one is no [longer with us], and two still left.’

Besides ANIMACY restrictions, it seems that the use of pronominal numerals denotes more specification (e.g. *dvox* ‘the two’) than collectives (e.g. *dvoe* ‘some two’). That is to say, pronominals would imply a greater level of individuation than collectives, but probably smaller than cardinals.

4.4.3.1.4. BCMS collective numeral substantives

Bosnian-Croatian-Montenegrin-Serbian (henceforth, BCMS) has developed a special category of collective numerals. Kim (2009) gives the following definition:

“[...] collective numeral substantives express a special meaning of the collectivity as a whole and a specification of gender, i.e., the members of a group are all males, but not mixed genders” [...]

The collective numeral substantives (e.g., *dvojica*, *trojica*, *petorica*, etc.) differ from the collective numerals (e.g., *dvoje*, *troje*, *petoro*, etc.) in that the former are nouns possessing their own ϕ -features, while the latter are modifiers which do not have their own ϕ -features. Furthermore, the collective substantives merge as the head of NP, while the collective numerals merge as the specifier of NP” (Kim 2009: 111-112).

According to Kim (2009: 120) collective numeral substantives (COLS) not only have their own φ -features, but they also have CASE (whereas collective numerals are most of the time indeclinable). For these reasons he considers that “[c]ollective numeral substantives are noun-like numerals” (*ibid*). Moreover, both BCMS collective numeral substantives and WP pronominal numerals are exclusively used with HUMAN nouns. Certainly, those properties sound closely related to what has been discussed about pronominal numerals in West Polesian, but we should be careful in comparing like with like.

BCMS collective numeral substantives (COLS) ostensibly look like pronouns. See, for example (55) a., where the numeral stands on its own (without specifying an NP). Nevertheless, most often collective numeral substantives in BCMS precede a noun as in (55) b.

(55) a. **BCMS** (In Kim (2009: 120) [modified])

Dvojica	su	doš-l-a	kući.
two(COLS).NOM	BE.3.PL	come-PST-F.SG	home

‘Two (males) came home.’

b. Kratk-a prič-a o **dvojici** star-ih prijatelj-a.
short-NOM.SG story-NOM.SG about two(COLS).LOC old-GEN.PL friend-GEN.PL

‘A short story about two old friends.’

We have already seen fin examples like (34) a.ⁱⁱⁱ and (52) a.ⁱⁱ that in West Polesian a construction like (55) b. (i.e. numeral + noun) is ungrammatical with a pronominal numeral.

Finally, although it is of less relevance, collective numeral substantives are only available for MASCULINES in BCMS, whereas WP pronominal numerals can be used with any noun denoting HUMANS. “They are feminine singular nouns in *-ica*, but ironically they specify a group whose members are all males” (Kim 2009: 111). In fact, collective numeral substantives are a solution or compensation to refer to an exclusively-male group, since regular collectives in BCMS imply a mixed-gender crowd.⁹² In sum, whilst the main function of collective numeral substantives is to refer to an all-male group of people as a whole, and they can seldom be used as pronominals (as a minor function), they have not developed a morphologically distinct class for that function, like in West Polesian.

4.4.3.1.5. Status numerals

In many Slavonic languages it is common to find nominalised forms of the cardinal numerals (STAT), which are commonly used in a metonymic way. Some examples of the most frequent uses are for referring to the number of a bus line or the name of a card (e.g. ‘you put down a five’). Leko (2009: 83) gives (56) a-e. as all meaning ‘Five presidents arrived’ in BCMS.

⁹² “Collective numerals designate the mixed gender specification of a group, that is, the heterogeneous combination of genders: male(s) and female(s). A group of females is modified by cardinal numerals, e.g., *tri prijateljice* ‘three female friends’ and *pet studentica* ‘five female students’ etc. Collective numerals usually do not decline after prepositions, but in formal contexts they can decline [...]” (Kim 2009: 113).

BCMS (In Leko (2009: 83)[modified])

- (56) a. Pet predsjednik-a je stig-l-o
 five(CARD) president-GEN.PL BE.3SG arrive-PST-N.SG
- b. Petorica predsjednik-a su stig-l-a
 five(COLS) president-GEN.PL BE.3PL arrive.PST-N.PL
- c. Petoro predsjednik-a je stig-l-o
 five(COLL) president-GEN.PL BE.3SG arrive-PST-N.SG
- d. **Petorka** je stig-l-a
 five(STAT.ANIMATE) BE.3SG arrive.PST-F.SG
- e. ***Petica** je stig-l-a
 five(STAT.INAMATE) BE.3SG arrive-PST.F.SG

‘[The] five [presidents] arrived.’

Examples (56) d. and (56) e. are instances of status numerals (STAT). Although functionally *petorka* is close to fulfilling the role of a pronoun in that context, the proper translation would be ‘the quintet’ (a highly nominalised form, to denote a well-known group of people), rather than ‘the five of them’. Yet, status numerals (STAT) are at odds with ANIMATES most of the time (i.e. unlike pronominals). Interestingly, Leko (2009) says that the form *petica* is not allowed (56) e., because it cannot be used to denote a group of humans. In any case, the primary function of status numerals is not to replace any NP (they are the noun *per se*); and in the narrowest sense, they do not quantify. For example, using the form (RU) *pjatërka* (from ‘five’) for a bus line does not specify the amount of bus lines (there may be more than five). In this respect, status numerals are close to ordinal numerals (§4.5.1.).

4.4.3.2. Pronominal numerals and other European languages

It is true that certain European languages allow the transformation of their numerals from pure Spec or Quantification into pronouns. There are some languages, such as Catalan or Ancient Greek, in which the numerals are pronominalized just by adding an article; e.g. *πέντε* ‘five’ vs. *οἱ πέντε* ‘the five’ (Dobrovský 1822 (in Viellard (2011)).⁹³ Other languages go just a little bit further. In German, the pronominalized numerals, apart from being preceded by an article, are capitalised in the orthography according to Suprun (1961); e.g. *sechs* ‘six’ vs *die Sechs* ‘the six’. Basque pronominalisation of numerals contains slightly more morphology than the others.

Basque

- (57) Bost ume etorri d-ira // d-a.
 five child.ABS come be.PRS-3PL be.PRS-3SG

Bost-ak elebidun-ak d-ira // (*d-a).
 five-ART.ABS.PL bilingual-ABS.PL be.PRS-3PL be.PRS-3SG

‘Five children have come. The five of them are bilingual.’

Nevertheless, the addition of the (DETERMINED) PLURAL ARTICLE (which in this case also marks ABSOLUTIVE) could have been added to another nominal constituent. Thus, the numeral does not present a dedicated morphological strategy. Now, it is important to notice one of the special effects of the pronominalisation of numerals. Whilst in the first sentence (57) the verb can stand either in the SINGULAR or PLURAL (*etorri da/dira*

⁹³ Many thanks to Helen Sims-Williams (p.c.) for confirming this.

‘have come’) without altering the meaning, which is a common phenomenon with numerals in Basque, in the second sentence (with the nominalised form of the numeral) it is ungrammatical to have a verb in the SINGULAR.

Up to now, I have not been able to find any language in which pronominal numerals have been attested as a distinct morphological class. Moreover, some of the most common sources for reference about typology and numerals, such as WALS (Dryer & Haspelmath 2013) or Corbett (2000), make no mention of it. Hence, the emergence of a morphologically dedicated class for pronominal numerals in West Polesian is not only an innovation within the Slavonic family, but apparently also in the European context, at least.

4.4.4. Summary

In sum, I have stated that West Polesian pronominal numerals embody a morphologically dedicated class. They have a double function; they replace an NP whilst also specifying the number of entities (humans). They are morphosemantically limited to numbers from ‘two’ to ‘ten’, and they can exclusively be used with nouns denoting humans. There are similar phenomena in other languages, but so far, I have not been able to identify any other language described as having such a particular class of numerals (at least in Europe).

4.5. Other numerals

4.5.1. Ordinals

As in any other Slavonic language, West Polesian has ordinal numerals. They do not display any property that we could not expect from other Slavonic or European languages. For this reason, I will describe them very briefly.

Ordinal numerals are morphologically adjectives, with the particularity of having a numerical value, which allows to order or arrange the NPs they qualify.

- (58) (T2.1 00:02)
jak id-aʃ u lis **parʃ-i** raz u fiód-u [...]
when go-2SG to forest.ACC.SG first-NOM.SG.M time.NOM.SG in year-LOC.SG
'When you go to the forest for the first time in the year [...]'.

Their morphological structure is almost predictably derived from cardinal numerals adding the suffix **-t(j)**[M]/**-ta**[F]/**-tae**[N]/**-te**[PL].⁹⁴ As in most (if not all) European languages the first numerals in the scale have suppletive stems (see more on suppletion in (Chapter 6)). The only two higher numerals (higher than 'ten') that escape the regularity are 'sorok > soroko'vi (forty > fortieth); and sto > 'sotni (a hundred > hundredth),⁹⁵ both inherited from Common Slavonic.

In Table 11, I present some of the most common ordinals. They all inflect like adjectives (i.e. they are available for all **GENDERS** and **NUMBERS**) but I have only indicated this with the first numeral, the rest take the same suffixes.

⁹⁴ Except for 'first', 'second', 'seventh', 'eight', 'fortieth' and 'hundredth'.

⁹⁵ Rather than suppletion there is some truncation of the stem.

Table 11 Main ordinal numerals in West Polesian

Ordinal	Gloss	Ordinal	Gloss
¹ <i>perɸ-ɪ</i> [M] <i>-a(ja)</i> [F] <i>-ae</i> [N] <i>ɪa/ɪj</i> [PL]	‘first’	<i>o¹ɖnatsɪɹɪ</i>	‘eleventh’
¹ <i>druɸɪj</i> <i>-a(ja)</i> ...	‘second’	<i>ɖva¹tsatɹɪ</i>	‘twentieth’
¹ <i>tretɹɪ</i>	‘third’	<i>trɪ¹tsetɹɪ</i>	‘thirtieth’
<i>tɕatj¹vjortɹɪ</i>	‘fourth’	<i>soroko¹vɪ</i>	‘fortieth’
¹ <i>pjetɹɪ/</i> ¹ <i>pjatrɪ</i>	‘fifth’	<i>pɪtjdɹ¹sjatrɪ</i>	‘fiftieth’
¹ <i>ɸostɹɪ</i>	‘sixth’	<i>ɸɪstdɹ¹sjatrɪ</i>	‘sixtieth’
<i>ɕjomɹɪ /sɪɖmɪj</i>	‘seventh’	<i>simdɹ¹sjatrɪ</i>	‘seventieth’
¹ <i>vosmɹɪ/</i> <i>vis¹mɹɪ</i>	‘eighth’	<i>visimdɹ¹sjatrɪ</i>	‘eightieth’
¹ <i>ɖevjatɹɪ/</i> ¹ <i>ɖrvjatɹɪ</i>	‘ninth’	<i>ɖrvɪ¹nostɹɪ</i>	‘ninetieth’
¹ <i>ɖɪsɕatrɪ</i>	‘tenth’	¹ <i>sotnɹɪ</i>	‘hundredth’

Being adjectives they agree in GENDER, NUMBER and CASE with their heads (Table 12).⁹⁶

Table 12 GENDER/NUMBER agreement with ordinal numerals

GENDER/NUMBER	Example	Gloss
a. MASCULINE	<i>tret-ɪ saraj</i>	‘(the) third barn’
b. FEMININE	<i>tret-a(ja) xata</i>	‘(the) third house’
c. NEUTER	<i>tret-ae jajtso</i>	‘(the) third egg’
d. PLURAL	<i>tret-ɪa nahavɪtsɹɪ</i>	‘(the) third (pair of) trousers’

⁹⁶ Examples reconstructed in order to create minimal pairs. Nevertheless, the evidence for it is based on an exhaustive observation of these patterns in the corpus.

Ordinal numerals can be nominalised to virtually function as pronouns; but again, pronominal numerals conform a different and morphologically dedicated category.

(59) (B20.8t 00:33)

sustrekaj-e tri anjolī “vi anjol-i, vi svjet-ie!
meet-3SG three.NOM angel.ADNM 2PL.NOM angel-NOM.PL 2PL.NOM holy-NOM.PL

tʃi batʃi-l-i moʃio sink-a?” **perʃ-i** kaʒ-e ja
PART see-PST-PL 1SG.POSS.ACC.SG son-ACC.SG first-NOM.SG.M say-3SG 1SG.NOM

ni batʃi-u”; **druʃi-i** kaʒ-e “ja tʃu-u b[i], ni
NEG see-PST.M.SG second-NOM.SG.M say-3SG 1SG.NOM hear-PST.M.SG PART NEG

tʃu-u”; **tret-i** kaʒ-e “ja sam tam bu-u”.
hear-PST.M.SG third-NOM.SG.M say-3SG 1SG.NOM REFL-NOM.SG.M there be-PST.M.SG

‘[She] meets three angels: “hail, holy angels! Did you see my son?” The first one says “I didn’t see him”; the second one says “no, I didn’t hear him”; [and] the third one says “I was actually there”.’

4.5.2. Fractions

Theoretically, the possibilities of morphologically deriving a fractional numeral from a cardinal are infinite (as cardinals are infinite). Nevertheless, the speakers only use a handful of them. When speakers need to refer to a less common fraction they tend to recur to circumlocutions, usually with *v* + CARDINAL (LOC) + *tʃastkax* ‘(lit.) in X pieces’. Sometimes it is unclear whether the fractional numeral speakers are using it as a loan from Russian (the language in which many were schooled), or whether it is a genuine West Polesian form. In any case, the most remarkable forms (with different variants) are *pov* ‘half’; *polo'vina/po'lova* ‘half’ (less frequent); and *povto'ra* ‘one and

a half'. Fractional numerals always govern GENITIVE (most frequently, SINGULAR).⁹⁷

Except for *pov*, the rest inflect and agree like FEMININE infl. class I nouns.

(60) a. (B10.6 02:12)

ʃtʃo tak-eje ʒe? **pov** fiolo'vi vʒe ni-ma!
Q.NOM this-NOM.SG.N PART half head.GEN.SG already NEG-HAVE

‘What is that? [You]’re already missing half of [your] head!’

b. (Z4.5 01:40) [Z1 speaking]

ja batʃi-l-a jak Tanja odleva-l-a **povtara**
1SG.NOM see-PST-F.SG how T.NOM pour-PST-F.SG one_and_a_half

lit-r-a, tʃi litrov-u bank-u i druʃi-eje **polovin-u**
litre-GEN.SG or one_litre-ACC.SG jar-ACC.SG and second-ACC.SG half-ACC.SG

‘I saw how Tanja was pouring half a litre... or perhaps a one-litre jar and an extra half.’

4.5.3. Distributives

There are two types of distributive constructions. The first one consists of simply using the suffix *po* + cardinal numeral. I have documented cardinal numerals being used in various cases (with no apparent discursive or sociolinguistic connotation), but most frequently they are used in NOMINATIVE (like in Bosnian (Leko 2009)), and less often in DATIVE (like in Russian). The examples (61) a., b. were produced by the same speaker (B20) in a time span of less than twenty seconds:

⁹⁷ Especially *pov* and *povtora*.

(61) a. (B14.9 00:12)

tam ʃeʃ [sic] 'patʃok, po 'deset ʃtuk
there six pack.GEN.PL for ten thing.GEN.PL

‘There are six packs there, each [with] ten things.’

b. (B14.9 00:25)

zd-am po ru'blj-u, po pɪ[t]dɪsjat ko'pɨjok
sell-1SG for ruble-DAT.SG for fifty kopek.GEN.PL

‘I sell each one for a [Belarusian] ruble, for fifty kopek (cents).’

c. (T5.2P 02:19)

zɪtʃɪ po korovaj-ovi
bake.PRF.INF for cake-DAT.SG

‘[Describing a task] to bake a cake for each of them.’

d. (B8.1 04:43)

po odn-ij bults-i dava-l-i [...] a otʃɪrɪd
for one-DAT.SG/just bun-DAT.SG give-PST-PL and queue.NOM.SG

bu-l-a, moʒe dvjestɪ tʃolovjek
be-PST-F.SG maybe two_hundred person.GRADNM

‘They used to give just a bun each, and there were perhaps two hundred people in the queue.’

The second option is to use a morphologically dedicated distributive numeral (DIS), which is only available for one number value (‘two’). The distributive *oba/obidva/oboje* means ‘both; the two of [them]’. In other Slavonic languages it is most often presented as following the same agreement as lower numerals, and thus taking ADNUMERATIVE.⁹⁸ Nevertheless, my observations do not allow me to conclude the same for West Polesian, so I cannot affirm that it allows the dedicated

⁹⁸ Because etymologically is derived from ‘two’ and it used to govern DUAL NUMBER (Suprun 1961).

ADNUMERATIVE form. Conversely, it can be said that it triggers what looks like GENITIVE SINGULAR nominal agreement and, in this instance, there seems to be more evidence to believe that it is probably a genuine GENITIVE SINGULAR form (see more on the debate of the forms in the next chapter, particularly in (§5.2.2.)). For these reasons, I decided to treat it separately from the proper (lower) cardinal numerals.

(62) a. (B7.6 03:06)

poduʃk-ı pokla-l-ı kob mlad-ije na poduʃk-ax si-l-i **oboje**
pillow-ACC.PL put-PST-PL COMP young-NOM.PL in pillow-LOC.PL sit-PST-PL DIS.NOM

‘[We] used to put down cushions, so that both the groom and the bride could sit.’

b. (P1.2.1 00:35)

tozje 'djetı je; dv-a 'xloptı, ženat-ije **oboje**
also child-NOM.PL be.PRS two-NOM.M boy.ADNM married-NOM.PL DIS.NOM

‘[I] also have children: two boys, both married’.

c. (Z4.el)

oba dzjıd-a moj-ı žıv-ıja
DIS.NOM uncle-GEN.SG 1SG.POSS-NOM.PL alive-NOM.PL

‘Both of my uncles are alive’.

4.5.4. Quasi-adverbial numerals

Quasi-adverbial numerals are also present in other Slavonic languages, at least in the Eastern Slavonic branch, as I have already mentioned in (§4.4.3.1.1.). They have not been the focus of my research, so more data would be required to contrast the information and extract more solid conclusions. Impressionistically, it is likely that the same morphosemantic limitation for pronominals and collectives applies here; i.e. being only available for numerals from ‘two’ to ‘ten’; although I have never documented any form

higher than ‘four’, and forms higher than ‘five’ are certainly rare in everyday speech. For example, (49) and (50) a.ⁱⁱⁱ (*supra*) are examples of this class.

(49) (B20.17 01:02)

i	lud-i	boja-l-i-sa	tʃerez	nɪfio	i-ti,
and	person-NOM.PL	scare-PST-PL-REFL	through	3SG.GEN.M	go-INF

obizatelno	ʃjob	udvox	tʃi	vtrox
necessarily	COMPL	two.QADV	or	three.QADV

‘People were afraid of crossing it [a haunted bridge], it was necessary [to do it] together with two or three people.’

Semantically, they are close in meaning to pronominal numerals and less so to collective numerals. Morphologically they are also close to pronominal numerals. In fact, the scanty collected data suggests that their form can be predicted (or parasitically derived) from pronominal numerals with the addition of the prefix *v/v-* (Table 13):

Table 13 Comparison of pronominals and quasi-adverbials

Pronominal	Quasi-adverbial	Gloss
dvox	vdvox	‘both together’
trox	vtrox	‘all three together’

In spite of the similarities with other numerals, there are also some syntactic differences. Quasi-adverbials often appear with a pronoun in NOM, whereas collectives always require the pronoun to be in ACC/GEN, and pronominals do not allow any other pronoun (given their idiosyncrasy) (63).

(63) a. ^I	mi	vdvox	a. ^{II}	*nas	vdvox
	1PL.NOM	two(QADV)		1PL.ACC/GEN	two(QADV)
	‘Both of us together.’				

4.6. Summary

In this chapter I have introduced the complex world of numerals in Slavonic and West Polesian. West Polesian has all the classes of numerals which other members of the Eastern Slavonic family have, but in addition to this, it has also developed a special category which I have called PRONOMINAL NUMERALS. This rarity has not been documented in any other Slavonic or even European language (to my knowledge), which opens a field for further research. Besides this phenomenon, I have shown the different classes of cardinals, what their effects are on the syntax, and how word order can also affect their meaning as it happens with postnominal numerals. Given that the system of cardinals presents some challenges for morphosyntactic agreement, I have shown that speakers can replace them by collective numerals. Nevertheless, collective numerals present some syntactic restrictions, particularly regarding ANIMACY (which are even more exaggerated with pronominal numerals). This has led me to reconsider introducing an additional level in the West Polesian Animacy Hierarchy: NON-HUMAN ANIMATES vs. HUMANS. I have ended up by describing some ‘minor classes’ (admitting this is inaccurate for ordinals). Nevertheless, none of these has any properties that cannot be found in any Eastern Slavonic variety.

Chapter 5

Adnumerative forms

Adnumerative forms are special forms that nouns take when followed by a numeral or quantifier, but which cannot be used elsewhere or stand on their own. In this chapter I analyse the adnumerative forms in West Polesian, particularly focusing on the (LOWER) ADNUMERATIVE. Firstly, I introduce the numeral phrases and adnumerative forms in Slavonic and I formulate my main question: What is the morphosyntactic nature of the ADNUMERATIVE? (§5.1.) I give examples of adnumerative forms in other languages (§5.1.2.), which also share many similarities with the West Polesian ADNUMERATIVE, although the latter one probably has the most robust adnumerative documented so far. Secondly, I explain the etymology of this form (§5.2.1.) and show its peculiar morphological (§5.2.2.) and syntactic (§5.2.3.) properties. Thirdly, I discuss whether there could be a GREATER ADNUMERATIVE form emerging in West Polesian (§5.3.), and identify the arguments both for (§5.3.2.) and against (§5.3.3.) such a hypothesis. Fourthly, I provide a brief overview of the existing analyses for closely related structures in Russian (§5.4.) and finally, I analyse the morphosyntactic nature of the adnumerative from a typological perspective and propose various analyses, by taking a canonical approach (CT) (§5.5.). I end with a summary of the chapter and the conclusions that can be drawn from it (§5.6.).

5.1. Introduction

In different languages, phrases containing numerals (henceforth, numeral phrases or NumPs) higher than ‘one’ govern different **NUMBER** forms. We would expect from languages like English to find **PLURAL** after two (64) a-c., but very so often we also find **SINGULAR**, as we can see from the examples in Hungarian (Kenesei et al. 1998, Rounds 2001) (64) d-f.:

English	Hungarian (Kenesei et al. 1998)
(64) a. <i>book</i>	d. <i>könyv</i> book
b. <i>book-s</i>	e. <i>könyv-ek</i> book-PL
c. <i>two book-s</i>	f. <i>két könyv</i> two book

Interestingly, some languages, such as the Slavonic family, have both **SINGULAR** and **PLURAL**. For example, Russian nouns headed by higher numerals (‘five’ to ‘twenty’) take **GENITIVE PLURAL** (henceforth, **GEN PL**), whereas lower cardinal numerals (‘two’ to ‘four’) are often described as governing what at least resembles **GENITIVE SINGULAR** (hereafter, **GEN SG**). However, in a NumP with a lower numeral, the adjective has a **PLURAL** form resulting in an apparent mismatch on the **NUMBER** values on the adjective and the noun it modifies (65) a-d.

Russian

- (65) a. (odin) xoroš-ij brat
 one.NOM.SG.M good-NOM.SG brother.NOM.SG
 ‘(One) good brother.’
- b. xoroš-ie bratj-a
 good-NOM.PL brother-NOM.PL
 ‘Good brothers.’
- c. pjat xoroš-ix bratj-ev
 five.NOM good-GEN.PL brother-GEN.PL
 ‘Five good brothers.’
- d. tri xoroš-ix brat-a
 three.NOM good-GEN.PL? brother-GEN.SG?
 ‘Three good brothers.’

In the case of Russian, there is a substantial amount of literature describing the phenomena around NumPs (numeral phrases), particularly the ones headed by lower cardinal numerals. Just to mention a few: (Akiner 1983, Babby 1987, Corbett 1978a, 1978c, 1983, Madariaga & Igartua 2017, Mel’čuk 1985, Rappaport 2002, Zaliznjak 1973, Žolobov 2003).

However, the phenomenon of NumPs in West Polesian brings the discussion to a different level. In Russian, a big part of the debate or mystery is how we explain the presence of the GEN SG with lower numerals (if we even believe it is a pure GEN SG). Now, data from recent fieldwork reveals that West Polesian uses an especially dedicated counting form, known as NUMERATIVE or ADNUMERATIVE (ADNM) for NumPs governed by lower (cardinal) numerals. Based on the descriptions in (Borsley et al. 2007, Corbett 2012, Nurmio & Willis 2016, Sims-Williams 1979), (forthcoming), I propose the following definition for the ADNUMERATIVE (ADNM):

ADNUMERATIVE (a.k.a. NUMERATIVE):

An inflectional form that nouns (or constituents of an NP) take when they appear in conjunction with a numeral (or less frequently a quantifier). This value can have a morphophonologically dedicated form for some nouns (or parts of the NP).

For the present work, I want to devote most of the attention to those instances where the West Polesian ADNUMERATIVE (ADNM) has a **unique or dedicated form**, because this is one of the phenomena that makes West Polesian special cross-linguistically.

According to Greville G. Corbett (p.c.), a broad definition of the ADNUMERATIVE (such as the one above) can classify as ADNUMERATIVE even the form of *book* in English *two books*, because there is a numeral phrase and there is an inflectional change in the noun (i.e. *book* is taking PLURAL, as result of *two*). But the form *books* is syntactically and semantically autonomous (i.e. it can stand on its own and it has meaning when it appears by itself). Since all these ‘NON-NOM SG forms after a numeral/quantifier’ are far too common cross-linguistically, it is not very practical to refer to them as ADNUMERATIVE forms. Therefore, I narrow down the focus to the *ad hoc* created inflectional forms, which are a far more marginal phenomenon in the world’s languages.

West Polesian

- | | | | | | |
|---------|------------------|------------------|----|------------------|------------------|
| (66) a. | (odɪn) | 'doxtar | c. | pjet | doxta'rɪv |
| | one.NOM.M | doctor(M).NOM.SG | | five.NOM | doctor(M).GEN PL |
| | '(One) doctor.' | | | 'Five doctors.' | |
| b. | doxta'rɪ | | d. | tri | 'doxtarɪ |
| | doctor(M).NOM.PL | | | three.NOM | doctor(M).ADNM |
| | 'Doctors.' | | | 'Three doctors.' | |

In (66) d. there is an example of a West Polesian ADNUMERATIVE (ADNM). Note that since the ADNUMERATIVE is realised in the entire form of the noun, i.e. segmentally and suprasegmentally (as I will explain in (§5.2.2.), I will not segment morphologically the glosses of nouns in these specific syntactic contexts.

That is to say, there is some syntactic ‘oddity’ in Russian NumPs containing lower (cardinal) numerals, whereas the morphology seems quite normal (i.e. it is believed to take normal GEN SG). Conversely, NumPs containing lower numerals in West Polesian are not only syntactically complicated, but morphologically they are also very unusual.

Let us have a look at the paradigm of the noun *traxtor* (M) ‘tractor’ in Table 14.⁹⁹

Table 14 Possible representations of the paradigm of *traxtor* (M) ‘tractor’

[a]	SINGULAR	PLURAL
NOM	^h traxtor	traxto ^h ri
GEN	^h traxtora	traxto ^h r-iv -ej
ADNM	^h traxtori	

Table 14 a.: The ADNUMERATIVE as a CASE value.

[b]	SINGULAR	PLURAL	ADNUMERATIVE
NOM	^h traxtor	traxto ^h ri	^h traxtori
GEN	^h traxtora	traxto ^h r-iv -ej	traxto ^h r-iv -ej

Table 14 b.: The ADNUMERATIVE as a NUMBER value.

After a glance at the data, an important question arises: Is the ADNUMERATIVE a CASE or a NUMBER value? The ADNUMERATIVE is used where we would otherwise expect the NOM PL (or at least GEN SG, like in Russian and Polish). So, is it a different NUMBER value of the NOMINATIVE (H1), or is it a special CASE value (H2)? In other terms, how are we going to analyse the data in Table 14: as (a.), a CASE; or (b.), as a NUMBER value?

⁹⁹ Given that the ADNUMERATIVE only appears in DIRECT CASES I only show the relevant cells.

However, since the ADNUMERATIVE displays traits of both CASE and NUMBER, but not fully (or very ‘non-canonically’, using Corbett’s (2012) terminology), can we treat the ADNUMERATIVE as **a feature on its own** (H3)? What do we expect to find if we believe so? Or, could the ADNUMERATIVE be a combination of two features (i.e. a **hybrid feature**) (H4)? I will try to answer to this question in this chapter.

The ADNUMERATIVE is cross-linguistically very rare. In this chapter, I show that West Polesian ADNUMERATIVE could be the most morphologically robust in the Slavonic family, and probably of all the documented languages with adnumeratives that I have been able to analyse. Thus, I will use data from West Polesian in order to shed some light on a cross-linguistically uncommon phenomenon. After analysing the etymology, the morphophonological form of the ADNUMERATIVE and the syntactic behaviour of NumPs, I address the question of its morphosyntactic nature by providing arguments for and against different analyses.

5.1.1. Methodological remarks

I have tried using natural speech from free texts for my analyses as much as possible. However, I arrived at a point where I was collecting lots of noise (as NumPs are not very prominent in the texts) and I was initially confused by the considerable inter- and intra-speaker variation of forms. For these reasons, I was in need of using prompts and sometimes even semi-direct elicitation.¹⁰⁰ In terms of location, it is true that the ADNUMERATIVE is present in every West Polesian variety covered in this work.

¹⁰⁰ Part of the parameters to develop the questionnaire on the syntactic behaviour have been extracted from (Shevelov 1963).

Nevertheless, amongst all the villages surveyed, the ADNUMERATIVE is particularly robust in Bahdanaŭka (the Pinsk district). Most of the analyses and examples here are based on that variety, B6 being my main language assistant and her son B9, who in most interviews stayed in the background giving corrections and instructions (and to whom I am also endlessly thankful).

5.1.2. Cross-linguistic overview

As noted earlier, the adnumerative (as a dedicated form) is a rare phenomenon cross-linguistically. I have only been able to find instances of adnumeratives in the Indo-European macro-family, concentrated in three families: Celtic, Slavonic and Indo-Iranian.¹⁰¹

5.1.2.1. Indo-Iranian

The Indo-Iranian family has some documented instances of robust adnumeratives. Amongst the Middle Iranian languages with an adnumerative, Sims-Williams (forthcoming) mentions **Sogdian** and **Choresmian**. Sims-Williams (1979) suggests that part of the cause of the development of adnumerative forms in Sogdian may have been the existing etymological ambiguity between the DUAL and the PLURAL suffixes. It is interesting to note that the ADNUMERATIVE was restricted to direct cases (NOMINATIVE-ACCUSATIVE) for all, if not the vast majority of nouns (and exclusively for nouns). And finally, the other important condition which characterised the Sogdian

¹⁰¹ Standard Basque uses a suppletive form for the noun ‘person’ ([ADNM] *lagun*; [SG/PL] *pertson-*), whenever it is headed by a numeral. But since it is a single example and it is mixed with suppletion, I have not included this in the survey.

ADNUMERATIVE was that in order to appear, the noun had to follow the numeral immediately (and its use was obligatory in these constructions).

The only other Middle Iranian language having an attested ADNUMERATIVE form is Choresmian, according to Sims-Williams (forthcoming). Sims-Williams suggests that the Choresmian adnumerative was also probably the result of an eroded dual. This form was at least optionally used with numerals higher than ‘four’ (although most examples include forms with lower numerals) and triggered PLURAL agreement on the adjectives. The morphosyntactic strategy employed to create this form consisted of using the (former) DUAL article -y, which was syncretic with the FEMININE article.

Within the new Iranian languages with an ADNUMERATIVE, Sims-Williams (forthcoming) mentions **Pashto**, **Parachi** and **Ossetic**, although probably the most interesting data comes from Ossetic. In Ossetic, the ADNUMERATIVE is only used in NOMINATIVE, although in contrast with Sogdian and other languages compared so far, the ADNUMERATIVE is not lost (i.e. replaced by a regular PLURAL form) when another constituent (usually, an adjective) appears between the numeral and the noun. The Ossetic ADNUMERATIVE concerns not only nouns governed by the numerals from ‘two’ to ‘four’ (as in Sogdian, Russian or Ukrainian), but also higher numerals and quantifiers (as in Bulgarian, Pashto or Parachi).

5.1.2.2. Celtic

According to Nurmio & Willis (2016), most of the nouns with an adnumerative in the Celtic family belong to two very specific semantic fields: kinship terms and units of

time (most commonly, ‘year’ and ‘day’).¹⁰² The origin of such a form is in the collapse (or deterioration) of a class of plurals, which end up being replaced by a new plural.

According to Nurmio & Willis (2016: 305), **Scottish Gaelic** retains a “numeral dual” (i.e. etymologically a DUAL, which synchronically is an ADNUMERATIVE, but it can only appear when the noun is headed by ‘two’). Yet it has only been preserved for a small group of FEMININE nouns.

The ADNUMERATIVE was once more widespread in **Middle Welsh** (according to Nurmio & Willis (2016)). For example, the noun for ‘brother’ had a special form after certain numerals: [one] *brawd*; [three] *broder*; [PL] *brodyr* ‘brother(s)’. **Modern Welsh** has only preserved one instance of adnumeral with the noun [SG] *blwyddyn* ‘year’ (Arwyn Watkins 1993, Borsley et al. 2007), which according to Arwyn Watkins (1993: 311) “*blynedd* must derive from an obsolete variant plural”.

When it comes to **Modern Irish**, Nurmio & Willis (2016) says that the adnumeral has only been retained for the noun ‘year’: [SG] *bliain*; [PL] *blianta*; [3-10] *bliana* ‘year(s)’.

5.1.2.3. Slavonic

Common Slavonic had a DUAL NUMBER value. There are some vestiges of residual DUAL morphology replacing the regular PLURAL (especially with the obliques) in many contemporary Slavonic languages; e.g. in Polish the noun [NOM.SG] *oko* ‘eye’ has an irregular paradigm [NOM.PL] *oczy* ‘eyes’; but [INSTR.PL] *oczyma* – ‘with (both) eyes’.

¹⁰² Although, according to Nurmio & Willis (2016: 305), Scottish Gaelic has an adnumeral for ‘foot’ (body part), which suggests that it may apply to semantically ‘natural pairs’ as well.

However, in most contemporary Slavonic languages the dual has been significantly eroded (except for Upper and Lower Sorbian, and Slovene which do preserve dual morphosyntax). The erosion of the Common Slavonic DUAL NUMBER has left a typologically peculiar trace in the counting systems of these languages. But I will deal with this in further detail in (§5.2.1.). So, for now, let us concentrate on the best-described instances of adnumeratives in Eastern and Southern Slavonic languages.

Looking at the Eastern Slavonic subfamily, **Standard Belarusian (BLM)** and **Standard Ukrainian (ULM)** are the best endowed in adnumerative forms (besides West Polesian) (Akiner 1983, Mayer 1971, Mayo 1976, Pugh & Press 1999). As far as I have been able to observe from different descriptions, ULM probably has a more robust ADNUMERATIVE than BLM, although in both languages adnumerative forms are in the process of decay under the influence of hypercorrection (Mayer 1971, Shevelov 1963). That is to say, speakers replace the dedicated ADNUMERATIVE forms by the NOM PL form, which is “the official rule” (for nouns headed by lower numerals) which they can deduce from most nouns. Nevertheless, their dedicated adnumerative forms have rarely been properly described as such (more details in (§5.2.1.)).

When it comes to **Contemporary Standard Russian (CSR)**, only seven nouns have a dedicated ADNUMERATIVE form (Corbett 2001, 2008, Mel’čuk 1985, Zaliznjak 1973, 2002, Žolobov 2003). The adnumeratives segmentally match the GEN SG form, but with a stress shift. Xolodilova (2015) remarks that whilst the nouns *šag* ‘step’; *čas* ‘hour’ are indisputably used in the dedicated ADNUMERATIVE form (in the specific contexts), there are another four to five nouns whose (dedicated) adnumeratives are not available in every speakers’ idiolect (also note that all the nouns with dedicated adnumeratives

belong to inflectional class II).¹⁰³ Otherwise, the GENITIVE SINGULAR is traditionally believed to replace the dedicated ADNUMERATIVE in such contexts (see more on the debate of Russian NumPs in (§5.4.)). CSR dedicated ADNUMERATIVES are also a ‘delicate’ form in terms of syntax, since the noun must immediately follow the head (numeral) or else, the GEN SG appears (67).¹⁰⁴

Russian

- (67) a. dva ča'sa
two hour.ADNM
'Two hours.'
- b.^I dva dolg-ix 'časa
two long-GEN.PL hour.GEN.SG
- b.^{II} #dva dolg-ix ča'sa
two long-GEN.PL hour.ADNM
'Two long hours.'

Within the Southern Slavonic sub-family, there are two types of ADNUMERATIVE forms, which may have a common origin, but synchronically they are completely different. The first type of ADNUMERATIVE is found in **Bulgarian** and **Macedonian** (Scatton 2002). Nouns from a specific inflectional class (allegedly all MASCULINE) show an ADNUMERATIVE form when governed by a numeral or a quantifier. Thus, this

¹⁰³ More on inflectional classes in (§3.2.1.).

¹⁰⁴ Based on (RusCorpora 2017). The simple query “*dva/tri/četyre dolgix časa*” only provided three hits. With a more detailed query, that is to say, giving morphosyntactic information of each constituent, it did not provide any results if the noun *čas* ‘hour’ was searched in ADNUMERATIVE, whereas it provided three hits, if *čas* was searched as GENITIVE. The same experiment was replicated for “*dva/tri/četyre širokix šaga*” ‘two/three/four wide steps’ but it only provided one hit in GENITIVE (for ‘three’).

ADNUMERATIVE is not exclusively employed with numerals, like in the rest of the Slavonic family.¹⁰⁵

The second ADNUMERATIVE type comes from **Bosnian-Croatian-Montenegrin-Serbian** (henceforth, BCMS). According to Alexander's (2006) description, BCMS has a very robust ADNUMERATIVE (probably more than the one in West Polesian), which is used with LOWER NUMERALS. Nevertheless, most authors have traditionally described it as a straightforward GENITIVE SINGULAR form, and unfortunately, Alexander's description is too brief to extract significant data to compare with West Polesian. What makes BCMS ADNUMERATIVE so special is that not only is it available for nouns, but also for adjectives and some demonstratives (unlike the rest of languages covered here). Moreover, it even has an effect on PAST TENSE GENDER/NUMBER agreement for MASCULINE nouns and it also affects agreement on ordinary adjectives (including in the predicate) (Alexander 2006, Corbett 1983).

“This form [...] looks very similar to the genitive singular, but it is NOT identical with it. For masculine and neuter nouns and adjectives the ending is **-a**. For feminine adjectives, the counting form ending is **-e** (but without vowel length), and for nouns it is like Gsg., but again without vowel length. ALL modifiers, including pronominal adjectives such as **taj**, take the counting form” (Alexander 2006: 59).

¹⁰⁵ Excluding fractionals ('half' and 'one and a half') and the distributive 'both' (at least) in most Slavonic languages ADNUMERATIVES are not permitted unless the numeral heading the NumP is a cardinal one (and certainly not quantifiers).

This seems the closest analogy to the West Polesian ADNUMERATIVE: it is morphologically robust, available to nouns from all inflectional classes; but semantically restricted to numerals ‘two’ to ‘four’.

Genetically speaking, the Standard Ukrainian (ULM) ADNUMERATIVE is a close analogous to the West Polesian ADNUMERATIVE. In fact, excluding West Polesian and BCMS, ULM has the second strongest ADNUMERATIVE in the Slavonic family.

5.1.2.4. Summary

What all the languages surveyed have in common is that they have a dedicated form (the ADNUMERATIVE) which marks the presence of a numeral or a quantifier, but that it cannot stand on its own to mean ‘two to four’. Up to now, we have seen that instances of adnumerative are rather an exception than the general rule. Firstly, in both the Indo-Iranian and the Slavonic family the adnumerative emerged as a result of the erosion of a DUAL NUMBER (based on Akiner 1983, Corbett 2012, Madariaga & Igartua 2017, Mayer 1971, Mel’čuk 1985, Sims-Williams (forthcoming), Žolobov 2003) which has only survived in the DIRECT CASES. In contrast, in the case of the Celtic languages, the ADNUMERATIVE emerged as the result of a disintegration of a PLURAL class, which was gradually replaced by a simpler PLURAL (Nurmio & Willis 2016: 303).

Secondly, in most languages the ADNUMERATIVE is semantically restricted, it only appears with cardinal numerals lower than ‘ten’; and those languages which allow the use of the ADNUMERATIVE with higher numerals also have the ADNUMERATIVE with other quantifiers.

Thirdly, in most languages the ADNUMERATIVE is only available for ‘MASCULINE’ and/or ‘NEUTER’ inflectional classes; and most often these are marginal inflectional (sub) classes. At first glance, the fact that nouns with an ADNUMERATIVE form belong to a marginal inflectional class seems to have contributed towards the fossilisation of the eroded DUAL (or NUMBER value) and the emergence of the ADNUMERATIVE. However, data from West Polesian shows that the ADNUMERATIVE is all over the noun system, which makes such an assumption harder to prove (i.e. did the ADNUMERATIVE extend by analogy from a small group, or did it just simply replace the DUAL?).

In the case of other languages like Bulgarian or Pashto (and perhaps Ossetic), the inflectional class which contains nouns with dedicated adnumerative forms is much larger (‘MASCULINE’ nouns ending in a consonant). But in contrast, in these languages the ADNUMERATIVE is not exclusively used with lower numerals. In order to be coherent with the analysis (i.e. comparing ‘like’ with ‘like’) I will treat this form as closely related, but different from the type of ADNUMERATIVE under study. I will call this type GREATER ADNUMERATIVE, because it extends to any NUMBER or quantifier. Its sister form, the one under study here, will be the (LOWER) ADNUMERATIVE, which is strictly for numerals, and which must be under ‘ten’ (unless they are a derived form). For the sake of practicality, I will refer to the LOWER ADNUMERATIVE simply as ADNUMERATIVE from now on.

And fourthly, Nurmio & Willis (2016) point that ADNUMERATIVE forms are “diachronically unstable”,¹⁰⁶ which seems right in the case of the Celtic languages, and perhaps the Indo-Iranian family. However, this does not apply to Ossetic, according to Sims-Williams (forthcoming) or Slavonic. Data from the Slavonic family (especially BCMS and West Polesian) reveals that such a generalisation is inaccurate.¹⁰⁷

Thus, after looking at data from other languages, West Polesian has one of the most robust documented examples of an ADNUMERATIVE (if not the most), as I show in the next section (§5.2.). See a summary in Table 15.

¹⁰⁶ “[...] the numeratives typically emerge from the disintegration of a major category such as plural or dual, and they are diachronically unstable, liable ultimately to analogical elimination” (Nurmio & Willis 2016: 297).

¹⁰⁷ I have already said that traditionally it has been explained that NumPs headed by ‘three’ and ‘four’ governed NOM PL. However, based on data from Contemporary Ukrainian and Belarusian, I suspect that that form was actually an ADNUMERATIVE rather than a NOM PL. Such a hypothesis still needs to be further elaborated, but should it be confirmed, it would mean that the ADNUMERATIVE has survived much longer than Nurmio & Willis (2016) ‘pessimistic’ claim suggests.

Table 15 Summary of the cross-linguistic survey

	Indo-Iranian	Celtic	Slavonic (excluding WP)
More than ten nouns have an ADNUMERATIVE	Sogdian + Choresmian? Pashto + Ossetic +	Modern Welsh – Middle Welsh? Modern Irish – Scots Gaelic –	BLM & ULM? ¹⁰⁸ CSR – BG & MKD + BCMS + ?
Can appear with any numeral	Sogdian – Choresmian – Pashto + Ossetic +	Middle & Modern Welsh – Modern Irish – Scots Gaelic –	BLM, CSR & ULM – BG & MKD + BCMS –
Etymology	Eroded DUAL	Eroded PLURAL	Eroded DUAL
Available for quantifiers which are not numerals (excluding distributives and fractions)	Sogdian – Choresmian – Pashto + Ossetic +	None of them.	BLM, CSR & ULM – BG & MKD + BCMS –

¹⁰⁸ Although not properly described as such in the literature.

5.2. Properties of the West Polesian adnumerative

The West Polesian ADNUMERATIVE can be compared to a morphological and morphosyntactic ‘Frankenstein’s monster’. NumPs containing lower numerals show a complex syntactic structure; and the ADNUMERATIVE form of the nouns itself takes odds and ends from different parts of the paradigm as we are going to see now.

5.2.1. Etymology

Etymologically, we know that the ADNUMERATIVE is the result of the erosion of the Common Slavonic DUAL number (in NOM/ACC). This erosion affected most Slavonic languages (except for Slovene and Upper and Lower Sorbian; (Corbett 2000, Corbett 2012), which has left an even more complicated government and morphosyntactic system as a result.

In Common Slavonic the numeral ‘one’ governed SINGULAR; ‘two’, DUAL; ‘three’ and ‘four’ NOM PL; and higher numerals GEN PL (see Table 16) (Akiner 1983, Hurski 1972, Žolobov 2003, to mention a few). This ‘change of gears’ between lower and higher numerals used to be even more remarkable in the past. According to Suprun (1961) and Corbett (1983), in Old Church Slavonic the first four numerals had morphologically a more adjectival behaviour, in that they could agree in GENDER with the numeral, whereas higher numerals did not permit such a thing. Moreover, whilst ‘two’ required the verbs to take DUAL and ‘three’ and ‘four’ PLURAL, higher numerals had both SINGULAR and PLURAL available (for verbal agreement) “with the SINGULAR prevailing” (Corbett 1983: 236).

Table 16 CASE and NUMBER government in Common Slavonic
(Akiner 1983, Žolobov 2007)

	‘one’	‘two’	‘three’, ‘four’	‘five’ to ‘twenty’
Common Slavonic	NOM SG	NOM DUAL	NOM PL	GEN PL

Over time, the DUAL started a process of decay whilst the numeral ‘two’ ended up merging with ‘three’ and ‘four’ into a single class of numerals (in terms of morphosyntactic agreement). The NOM/ACC DUAL was (phonologically) syncretic with the GEN SG for two of the main inflectional classes (the **-o* stem (containing primarily MASCULINE nouns) and the **-i* stem) at least at some point, so that may have caused some Slavonic languages to develop an ADNUMERATIVE form on the basis of the GEN SG (e.g. Russian, Polish) (Žolobov and Krys’ko 2001, Žolobov 2003, 2007).¹⁰⁹ Others (at least Ukrainian and Belarusian) have taken NOM PL as the basis for their counting forms (Hurski 1972, Shevelov 1963).¹¹⁰

Traditionally, the forms that nouns take after numerals have been described as specific CASE/NUMBER values. This is given the similarities with other ‘well-behaved’ cells of the paradigm in contemporary Slavonic languages. See Table 17.

¹⁰⁹ Even though this has been the most commonly accepted position until recently, Madariaga & Igartua (2017) provide solid historical evidence that refutes this claim about Russian (see more in (§ 5.4.)).

¹¹⁰ In the broadest understanding of ADNUMERATIVE, any form other than NOM SG after a numeral would qualify as ADNUMERATIVE. However, I have previously specified that for the premises of this work, I only focus on the dedicated ADNUMERATIVE forms. When it comes to Belarusian and Ukrainian, they take NOM PL ‘by default’ which is different from the dedicated ADNUMERATIVE forms that they have also developed.

Table 17 NUMBER/CASE agreement (for nouns) in Eastern Slavonic and Polish

Language	'one'	LOWER NUMERALS			HIGHER NUMERALS
		CLASS-II	CLASS-I	CLASS-III	
Belarusian	NOM SG	NOM PL	NOM PL	NOM PL	GEN PL
			GEN SG	GEN SG	
Ukrainian	NOM SG	NOM PL	NOM PL	NOM PL	GEN PL
			GEN SG	GEN SG	
Polish	NOM SG	[VIRILE] NOM PL	GEN SG	GEN SG	GEN PL
		[rest] GEN SG			
Russian	NOM SG	GEN SG	GEN SG	GEN SG	GEN PL

Such simplifications may provide a good general description of how numerals govern nouns in those languages. However, already in Standard Belarusian (BLM), and in Standard Ukrainian (ULM) linguists often argue about which CASE/NUMBER form inflectional class I and III nouns take (i.e. whether it is GEN SG or NOM PL) (Akiner 1983, Hurski 1972, Lepešaŭ 2002, Mayer 1971, Mayo 1976, Piskunoŭ 2010, Pugh & Press 1999, Rusanovskij et al. 1986, Shevelov 1963, Vjarxoŭ 1961).¹¹¹ Some have proposed complex syntactic rules, which would be lexically specified, or at least inflection class

¹¹¹ Most authors avoid talking in depth about counting forms in BLM and ULM. Particularly when it comes to inflectional class I, many have provided confusing and complicated syntactic rules. Although it must be said in their favour that West Polesian inflectional paradigms are a lot more transparent than those of Standard Belarusian (as they allow more stress pattern combinations), which are a considerable aid for understanding what is happening below the surface. The most complete, but perhaps also confusing analysis of the counting forms that appear in BLM was carried out by Piskunoŭ (2010). In this article, the author makes a considerable corpus study trying to set some regularities or rules to determine which CASE/NUMBER form appears after a numeral. The author manages to create a few groups of CASE/NUMBER forms, but the irregularities and exceptions often outnumber these 'regular' nouns. Nevertheless, in the light of the data from West Polesian, there are good reasons to believe that a better analysis is needed.

sensitive, but which are very problematic for a “Morphology-free-Syntax” (Baerman et al. 2005) approach.¹¹²

In any case, these descriptions cannot give a good account of data from West Polesian because it adds a new element to the equation: the ADNUMERATIVE (ADNM). As I have shown in (§5.1.2.3.), West Polesian is not the only Slavonic variety with an adnumerative, but altogether with BCMS they both have the most morphologically robust adnumerative in the family.

It must be said, since DABM (1963) and Mackevič et al. (1964) certain Belarusian dialectologists (e.g. Hulickaja et al. 1992) have very concisely documented the existence of special forms of the nouns (especially for class I and III nouns) which was distinctive (or especially prominent) in south-western Belarusian varieties (including the area of West Polesian). Yet, most, if not all, works have hardly paid any attention to the ADNUMERATIVE. They claimed that a certain form was used in a certain variety, but no analysis was really provided (because that was not part of their plan). Some have made a bit superficial analyses which simply refer to it as a “remnant (of the) dual” (e.g. Hulickaja et al. 1992, Mackevič et al. 1964).

“Many linguists (Filin 1972: 30-31) relate the presence of the forms –è and –a in the dialects with the forms of the dual number, which were widely used in Slavonic languages. The loss of the paradigms of the dual number is a process that was in progression at different times and not uniformly in the territory of the Eastern Slavonic languages. This can be

¹¹² I.e. it would disprove the claim that numerals can govern different CASE/NUMBER forms in the noun depending on its inflectional class.

attested in ancient manuscripts as well as in contemporary dialects. This way, the feminine of such forms started to disappear earlier in the eastern part of middle-Russian dialects, whereas in western and Belarusian dialects such nouns are distinguished by a stronger stability of the dual number. Such stability is also characteristic of the neuter, which can be proved with examples [...] Thus, the extension and stability of the dual forms in Belarusian dialects can be seen as a morphological peculiarity in comparison with Russian dialects” (Hulickaja et al. 1992: 61-62 (in Roncero 2015)).

Yet, as we have already seen and I explain further in this chapter, any slightly deeper analysis shows that the ADNUMERATIVE is far from being a DUAL number value (synchronically).

In addition to this, it must be said that the data in DABM regarding constructions of lower numeral + noun do not always match the findings of my own (recent) fieldwork. However, probably my biggest objection to the quality of the work on this parameter (i.e. lower numerals + nouns) in the literature of Belarusian dialectology is related to the way DABM (the biggest reference for Belarusian dialectology) formulated the research question for feminine (i.e. inflectional class I) nouns headed by lower numerals. The authors of DABM documented under one single segmental form two different CASE values that are distinguished by prosody in certain varieties. For example, they use the noun *xatʲi* ‘houses’ as a test noun, which without further details, could be either [NOM PL] /xa'tʲ/ (at least in WP) or [GEN SG] /'xatʲ/ in West Polesian. That is to say, they obliterated the importance of the distinctive function of the stress in certain varieties (in West Polesian, there are more distinctions than in BLM), and as a result, the morphology is obscured.

5.2.2. Morphology

In order to be able to see the ADNUMERATIVE in the noun paradigms, it is necessary to recall the relevance of prosody (stress) in West Polesian. Stress is phonemic across the entire lexicon in West Polesian and often disambiguates segmentally identical forms in inflectional paradigms (see (§3.1.3.)). I should remind the reader that since CASE/NUMBER marking involves the whole stem (i.e. the inflection and the stress) I do not provide a morphological segmentation of nouns in numeral phrases in this chapter.

Parasitically, the morphophonological shape of the ADNUMERATIVE can be derived from the stem of the NON-DIRECT (or OBLIQUE) SINGULAR cells (or at least the GEN SG) + the suffix *-i*. As far as I have been able to observe, the suffix *-i* is commonly shared by all the inflectional classes. For that reason, the ADNUMERATIVE is only phonologically distinct for inflectional class II and III nouns with dynamic stress (69), (70). That is to say when the SINGULAR sub-paradigm and the PLURAL sub-paradigm follow different stress patterns (i.e. on the stem, or on the inflection). In order to reinforce the importance of the stress in the paradigms, besides the cell of the ADNUMERATIVE, also compare (69) b. with (69) d.

Nevertheless, the ADNUMERATIVE is syncretic with some forms of the paradigm (NOM PL, GEN SG and or LOC SG) when the stress falls in the same syllable. The infl. class I ADNUMERATIVE is always syncretic with the GEN SG, as they share the same inflectional suffix and type of stem/stress. And when the stress is static it is also syncretic with NOM PL for the same reasons ((71) *xlopets* (CLASS II) ‘boy’). Thus, the rules of formation of the ADNUMERATIVE are similar to these other (well-established) forms, and so adnumerative forms can end up looking the same *on the surface level*. Yet, the resulting

syncretism of the forms is merely accidental and caused by the position of the stress (and the identity of the inflection).

(68) *brat* (CLASS II) ‘brother’

- a. (odm) *brat*
one.NOM.M brother(M).NOM.SG
‘(One) brother.’
- b. *bra'tɪ*
brother(M).NOM.PL
‘Brothers.’
- c. *tri* **'bratɪ**
three.NOM brother(M).ADNM
‘Three brothers.’
- d. *'brata*
brother(M).GEN.SG
‘[The] brother’s [GEN SG].’

(69) *jajtso* (CLASS III) ‘egg’

- a. (odn-e) *jaj'tso*
one.NOM.N egg (N).NOM.SG
‘(One) egg.’
- b. *'jajtsa*
egg(N).NOM.PL
‘Eggs.’
- c. *tri* **jaj'tsɪ**
three.NOM egg(N).ADNM
‘Three eggs.’
- d. *jaj'tsa*
egg(n).GEN.SG
‘[The] egg’s [GEN SG].’

(70) *korova* (CLASS I) ‘cow’

- a. (odn-a) *ko'rova*
one-NOM.F cow(F).NOM.SG
‘(One) cow.’
- b. *koro'vɪ*
cow(F).NOM.PL
‘Cows.’
- c. *tri* **ko'rovɪ**
three.NOM cow(F).ADNM
‘Three cows.’
- d. *ko'rovɪ*
cow(F).GEN.SG
‘[The] cow’s [GEN SG].’

(71) *xlopets* (CLASS II) ‘boy’

- a. (odm) *'xlopets*
one.NOM.M boy(M).NOM.SG
‘(One) boy.’
- b. *'xloptsɪ*
boy(M).NOM.PL
‘Boys.’
- c. *tri* **'xloptsɪ**
three.NOM brother(M).ADNM
‘Three boys.’
- d. *'xloptsa*
brother(M).GEN.SG
‘[The] boy’s [GEN SG].’

In the NON-DIRECT CASES (i.e. all but NOM and ACC) the effect of the heading numerals is neutralised. Nouns which had a dedicated ADNUMERATIVE form when headed by ‘two’, ‘three’ and ‘four’ (in direct cases) share the inflectional cells with the plural sub-paradigm.

- (72) a. $si'ni$
 son.NOM.PL
 ‘Sons.’
- b. $dv-a$ $'sni$
 two-NOM.M son.ADNM
 ‘(Two) sons.’
- c. z $si' n-amɪ$
 PREP son-INST.PL
 ‘With (the) sons.’
- d. z $dv-uma$ $si' namɪ$
 with two.INST
 son.INST.PL/ADNM?
 ‘With two sons.’

This has an explanation if we look at the phenomenon in diachrony. We have said that the ADNUMERATIVE is the direct result of an eroded DUAL NUMBER. This Common Slavonic DUAL had fewer CASE distinctions than the SINGULAR and PLURAL subparadigms (although they were different from SINGULAR and PLURAL). According to Madariaga & Igartua (2017):

“[d]ual number had a separate inflectional paradigm that distinguished three forms (with syncretisms for nominative-accusative, genitive-locative, and dative-instrumental values) and exhibited allomorphic differences depending on inflectional class” (Madariaga & Igartua 2017: 105).

Thus, we can see that the only part of the DUAL NUMBER paradigm that has been passed on is the NOMINATIVE-ACCUSATIVE (i.e. DIRECT CASES).

Very frequently nouns that should take the morphophonologically dedicated ADNUMERATIVE can take other forms with (so far) no demonstrated semantic or discursive

distinction.¹¹³ Thus, there seems to be case/form competition. This might have proliferated over the past decades under the increasing influence of Russian and Polish (less so Standard Belarusian and Ukrainian) in the area I studied. Admittedly, there is evidence of NOM PL /GEN SG competition up to the nineteenth century in Russian and Ukrainian (Madariaga & Igartua 2017, Mayer 1971, Žolobov 2003). There is diatopic variation, in terms of the frequency of variation, but, surprisingly, the age or the level of bilingualism does not have a significant influence on the speakers' choices of one or another form. See in (73) all the forms proposed by B6 and her son B9 for 'two oak trees':

- (73) (B6, B9.el)
- | | | |
|-------------------|-----------|---------------|
| a. ^I | dv-a | 'duba |
| | two-NOM.M | oak(M).GEN.SG |
| a. ^{II} | dv-a | du'bɪ |
| | two-NOM.M | oak(M).NOM.PL |
| a. ^{III} | dv-a | 'dubɪ |
| | two-NOM.M | oak(M).ADNM |
- 'Two oak trees.'

5.2.2.1. Alternatives to the counting system:

Besides the constant variation of suffixes, which at least resembles other forms of the paradigm (i.e. the ones that are not the phonologically distinct ADNUMERATIVE), the other key difficulty for eliciting the ADNUMERATIVE has been the existence of a parallel counting system, which I have already briefly introduced in the previous chapter (§4.3.2.2.). Apart from the system consisting of CARDINAL NUMERAL + NOUN that I have

¹¹³ Amongst all the analysed texts I have only identified one instance in which the alternation seemed to be conditioned by the prosody (to match the rhyme). Example from a tale: [B20] *ʒɪ'li dva 'bratɪ; 'bidnɪ i ba'fɪatɪ* '[Once] there were two brothers: one [was] poor, and [the other one] rich'.

been describing,¹¹⁴ there is another (apparently emerging) system which consists of COLLECTIVE NUMERAL + NOUN. The advantage of this system is that the agreement of the rest of the constituents as well as the marking on the noun is more straightforward: always GEN PL.¹¹⁵ In the examples below (74) a.^I and a.^{II} compare the NumP ((74) a.^I) which is headed by a cardinal numeral to NumP ((74) a.^{II}) which is headed by a collective numeral. Both forms were proposed by B20 (Bahdanaŭka, male) as synonyms in this context.¹¹⁶

- (74) (B20.el)
- | | | | | | |
|-----------------|---------------|------------|------------------|-----------|--------------|
| a. ^I | dv-a | ko'ni | a. ^{II} | dvoje | 'konej |
| | two-NOM.M | horse.ADNM | | two(COLL) | horse.GEN.PL |
| | 'Two horses.' | | | | |

Standard Ukrainian is undergoing the same process according to Mayer (1971), but other than that, this relatively free use of collectives looks unacceptable in other Slavonic languages (e.g. Russian). Impressionistically, this phenomenon seems more prominent in men's speech than in women's. In terms of agreement, the 'collectivised' NumPs behave like regular NumPs headed by collective numerals (§4.3.3.).

5.2.3. Morphosyntactic behaviour of NumPs containing lower numerals

In this sub-section I explain the morphosyntactic behaviour of NumPs containing lower numerals, and thus, potentially containing dedicated adnumerative forms. First, I

¹¹⁴ And thus, taking the ADNUMERATIVE when the NumPs contain a lower numeral.

¹¹⁵ Or GREATER ADNUMERATIVE depending on the analysis.

¹¹⁶ More on ANIMACY and its restrictions on non-cardinal numerals in (§4.3.3.3.1.) and (§4.4.2.3.3.).

present the agreement of these NumPs at the level of the noun phrase (§5.2.3.1.). I present the different configurations that the insertion of adjectives allow and how they also complicate the agreement riddle. Then, I move into the particular challenges ANIMACY poses for NUMERACY or counting leading to conflicts which can resolve in different ways (§5.2.3.1.1.). Finally, I present the agreement pattern that phrases containing lower cardinal numerals trigger on verbs, which I suggest differs little from the verbal agreement triggered by higher cardinal numerals or collectives (§5.2.3.2.).

5.2.3.1. Agreement in the noun phrase (NP)

When the NP is governed by a lower numeral all the constituents (except for the noun in ADNUMERATIVE) take PLURAL, either NOM PL or GEN PL (or what looks like these CASE/NUMBER forms). The GEN PL is especially frequent when the constituent appears after the heading numeral. Though, as far as I have been able to see in the corpus the combination *[NOM PL] adjective + [ADNM] noun is not allowed (75).

(75) (P2.3.el)

- | | | | |
|----|----------------------|----------------------|----------------|
| a. | tri | lasnjovat- ix | ' sin-i |
| | three.NOM | lazy-GEN.PL? | son(M).ADNM |
| | 'Three lazy sons.' | | |
| b. | tri | lasnjovat- ix | ' sin-a |
| | three.NOM | lazy-GEN.PL? | son(M).GEN.SG |
| | 'Three lazy sons.' | | |
| c. | tri | zdorov- ie | si'n-i |
| | three.NOM | strong-NOM.PL | son(M).NOM.PL |
| | 'Three strong sons.' | | |

- d. *tri zdorov-**ie** 'sin-**i**
 three.NOM strong-NOM.PL son(M). ADNOM
 'Three strong sons.'

Other tests carried out with speakers also show a strong dislike for the combination of a noun in ADNUMERATIVE with an attributive adjective in NOM PL, even when the noun form is phonologically identical to the NOM PL (e.g. the noun [NOM SG] 'xlopets [NOM PL = ADNOM] 'xloptsɪ 'boy(s)).

If the adjective stands in a predicative function, the adjective will simply take PLURAL (NOM PL if it is in a DIRECT CASE) (76).

- (76) (B6.32.el)
 djɪd bofiat-**i**; jɪfo dv-a 'traxtorɪ
 uncle.NOM.SG rich-NOM.SG POSS.3SG.M two-NOM.M tractor.ADNOM

 susim nov-**i**je
 totally new-NOM.PL

 'That man (lit. 'uncle') is rich. His two tractors are brand-new.'

Conversely, if the adjective appears in a **prenumeral** position (as well as any other constituent) most often it takes NOM PL agreement. The best explanation for this seems that when a constituent is inside the NumP it is fully governed by the heading numeral, but once it is outside of it, it identifies the whole phrase as a PLURAL (or NON-SINGULAR), and hence, it is no longer syntactically bound to the constraints inside the phrase (77).¹¹⁷

¹¹⁷ NB: the noun in this example belongs to a different inflectional class from the previous example (76).

(77) Proposed sentence:

tʃatɪrɪ nɪvɪdomɪx mɪnɪ zənʃtʃɪnɪ vajʃlɪ v podjezd.

‘Four unknown (to me) women got into the train.’

(B6.el)

njak-ije	dv-i	nɪznakom-ije	ʒo'notʃɪnɪ
some-NOM.PL	two.NOM.F	unknown-NOM.PL	woman.ADNM/NOM.PL?

vojʃ-l-i	v	pojezd
enter-PST-PL	in	train.ACC.SG

‘(Some) two unknown women got into the train.’

This also applies to nouns from the infl. classes II and III. Compare (78) a. with (78) b.^I:

(78) (B9.el)

a.	tri	'bratɪ ¹¹⁸	pirɪjɪxa-l-i	v	Amerɪk-u
	three.NOM	brother.ADNM	move.PST.PL	to	America-ACC.SG

‘[The] three brothers moved to America.’

b. ^I	vs-i	tri	bra'tɪ	pirɪjɪxa-l-i	v	Amerɪk-u
	all-NOM.PL	three.NOM	brother.NOM.PL	move.PST-PL	to	America-ACC.SG

b. ^{II}	*vs-i	tri	'bratɪ	pirɪjɪxa-l-i	v	Amerɪk-u
	all-NOM.PL	three.NOM	brother.ADNM	move.PST-PL	to	America-ACC.SG

‘All three brothers moved to America.’

However, later on B6 approved several sentences, in which the prenumeral adjective *vsi* ‘all’ was in NOM PL and the noun in the NumP in ADNUMERATIVE (79):

¹¹⁸ Although I have also documented **tri bra'tɪ** [ADNM = NOM PL].

(79) (B6.27.el)

a.^I **vs-i** **tʃtɪrɪ** **'bratɪ** **zrobi-l-i** **sistr-i**
all-NOM.PL four.NOM brother.ADNM make.PRF-PST-PL sister-DAT.SG

visilj-e
wedding-ACC.SG

a.^{II} **vs-i** **tʃtɪrɪ** **bra'tɪ** [...]
all-NOM.PL four.NOM brother.NOM.PL

‘All four brothers made a wedding for [their] sister.’

b. **vs-i** **tʃtɪrɪ** **'sɪnɪ** **zafinu-l-i** **na vojnj-i** **u jji**
all-NOM.PL four.NOM son.ADNM die-PST-PL in war-LOC.SG in 3SG.GEN.F

‘All her four sons died in the war.’

The same rule (i.e. having a prenumeral modifier in NOM PL) also applies to NumPs containing higher numerals (80):

(80) (B6.27.el)

vs-i **sem** **bra't-ɪv** **ʒiv-utj** **v** **sel-i**
all-NOM.PL seven.NOM brother-GEN.PL live-3PL in village-LOC.SG

‘All the seven brothers live in the village.’

Corbett (1983) describes an analogous construction in Russian (with the numeral ‘five’):

“What is special about this is that the numeral [...] is treated as a modifier: it marks the boundary between modifiers showing full agreement and those in the nominative [...]” (Corbett 1983: 218).

The two exceptions I have found may be motivated because the modifier in both sentences is not a pure adjective, but more of a quantifier (which may trigger a partitive reading). They were both elicited from B9 (82).¹¹⁹

(81) (B9.el)

- a. **tsɪl-ix** tri dnji fuɭja-l-ɪ vesilj-e
 whole-GEN.PL three.ACC day.ADNM to_party-PST-PL wedding-ACC.SG

‘[People] used to party for three whole days at the wedding[s].’

- b. vin pi-v 'tsɪl-ix tri 'pljaʃkɪ
 3SG.NOM.M drink-PST.M whole-GEN.PL three.ACC bottle.ADNM

somo'fionk-ɪ

samogon-GEN.SG

‘He drank three full bottles of samogon.’

5.2.3.1.1. NP agreement and animacy

The fact that the NumPs sometimes behave like ‘islands’ (borrowing quite freely the metaphor from generativist terminology) or ‘shields’ to certain syntactic effects can be proved by ANIMACY. Let me illustrate this with some examples.

The first one (82) comes from an overheard sentence. I visited the church in Kamien with B22, a speaker from the neighbouring village during a cold day. As we were leaving, an old man who wanted to be hospitable said aloud:

¹¹⁹ Note that in Russian a construction like (81) b. would also require a GEN PL (instead of GEN SG/ADNM).

- (82) (Overheard 12/02/2017 in Kamien)
- | | | | | |
|------|----------|-----------------|------------|-----------------|
| nada | bra-tj | fiet-ix | dva | 'bratɪ |
| need | take-INF | this.ACC/GEN.PL | two.NOM?M | brother(M).ADNM |
- ‘It’s necessary to take these two brothers.’¹²⁰

The verb *bratj*/'*bratɪ* ‘to take’ requires an object to be in ACCUSATIVE.¹²¹ As in many Slavonic languages West Polesian ACCUSATIVE is a non-autonomous CASE (except for the Inflectional Class I, -a) (more on ANIMACY in (§3.4)). The way the ACCUSATIVE decides which form to take depends on ANIMACY. If the noun is ANIMATE, it takes an ACC = GEN form, or else, if INANIMATE, it takes the ACC = NOM form. In the sentence in (82) we see that the determiner has an ACC = GEN form (*fietɪx*), as it identifies the noun as ANIMATE. We would expect the noun or NP which is determining to be in ACC = GEN as well: either SINGULAR (*dva 'brata*), if *dva* does not take the ACC = GEN; or most likely PLURAL (*dvux brat-iv/-ov*). But none of this is happening: NUMERACY (i.e. quantification with cardinal numerals) trumps ANIMACY in this context (admitting that I have also recorded many utterances where ANIMACY is overriding NUMERACY). That is to say, the syntactic effect of ANIMACY cannot penetrate the ‘shield’ created by the numeral phrase. Thus, the syntactic structure of (82) is Figure 5:

¹²⁰ Meaning ‘someone should take care of (i.e. invite home) these two brothers’.

¹²¹ The proper West Polesian suffix for infinitives is -*ti* (i.e. '*bratɪ* which, by the way, is homophonous with the ADNUMERATIVE of *brat* ‘brother’). The fact that he used a more ‘Russian-like’ suffix may have been motivated by the sociolinguistic context (Protestant churches in Belarusian Western Polesie have most, if not all, of their liturgy in Russian (see (§1.2.3.)).

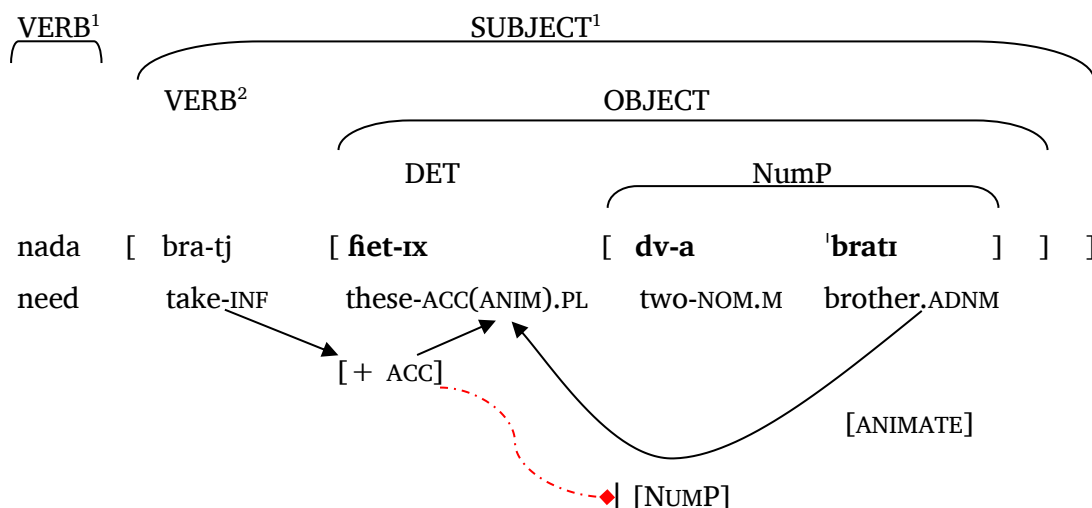


Figure 5: Syntactic structure of (82) (NumP blocking ANIMACY)

The sentence in (83) is another example of a numeral phrase blocking **ANIMACY** (in a free text). Note that the object of the sentence does not take the [ACC = GEN] form we would expect (the object is in [ACC = NOM]).

(83) (T11.6 00:25)

tut	dv-a	tʃolo'viki,	dv-a	'xloptsi
here	two-NOM.M	man.ADNM	two-NOM.M	boy.ADNM
zabi-l-o	moljni-j-a			
kill.PRF-PST-N.SG	lightning(N)-NOM.SG			

‘The lightning killed two men, two boys here.’

Some may argue that the neutralisation of **ANIMACY** in (83) is in reality caused by the topicalisation (or fronting) of ‘two men’. However, I have obtained more examples from elicitation, in which animacy distinctions dissipate in the presence of numerals (and nouns in ADNUMERATIVE). In neither of the sentences (84) a., b. does **ANIMACY** reach the last constituent in the chain: the noun. Notice that in (84) a. **ANIMACY** does not even reach the prenumeral modifier, which is outside of the NumP:

(84) (B6.30.el)

- a. ja batʃ-u ʃet-i dv-a bra'ti.
1SG.NOM see-PRS.1SG this-NOM/ACC.PL two-NOM.M brother.NOM.PL
'I see these two brothers.'

Even though B6 also said immediately:

- b. ja batʃi-l-a dv-a duʒ-ix bra'ti¹²²
1SG.NOM see-PST-F.SG two-NOM.M big-ACC/GEN?PL brother.NOM.PL
'I [female speaking] saw [the] two big brothers.'

In sum, when NUMERACY and ANIMACY meet in a phrase there is often a conflict which tends to be resolved in favour of the NumP. Having said this, it must also be remembered that the ADNUMERATIVE form does not distinguish between NOM and ACC, and so it can fulfil both syntactic roles. Given the ambiguity it allows, we would expect it to appear more often as a way of resolving conflicts, in favour of the NOM PL form, which is not necessarily the case in most of the examples I have shown here (or which I have collected). Being a syntactically 'delicate' form (i.e. it has to immediately follow the numeral in order 'to survive') it may be that ANIMACY also imposes a burden upon the dedicated ADNUMERATIVE form.

5.2.3.2. Verbal agreement

Lower numerals (and thus NumPs with the ADNUMERATIVE) trigger PLURAL agreements in the verb in the present tense (and derived forms) (85) a-d.:¹²³

¹²² She produced the second sentence in the PAST tense, although it does not have any effect on the syntax of the numeral phrase.

¹²³ When the FUTURE is used (usually for equations) then using SINGULAR seems to be the norm.

- (85) (B6.el)
- a. na dvor-i firaj-e odm xlopets
in outside-LOC.SG play-3SG one.NOM.M boy.NOM.SG
‘A boy is playing outside.’
- b. na dvor-i firaj-utj dv-a 'xlopts-i
in outside-LOC.SG play-3PL two-NOM.M boy.ADNM
‘Two boys are playing outside.’
- c. na dvor-i firaj-utj tʃtir-i 'xlopts-i
in outside-LOC.SG play-3PL four.NOM boy.ADNM
‘Four boys are playing outside.’
- d. na dvor-i firaj-utj 'xlopts-i
in outside-LOC.SG play-3PL boy.NOM.PL
‘(The) boys are playing outside.’

So far, I have only found one exception in the whole corpus (86). In a context where we would expect to find a noun in ADNUMERATIVE (instead of the GEN SG she used) the verb stands in SINGULAR.¹²⁴

- (86) (Z7.8 01:06)
- ja batʃi-l-a, ʃto tam tri malenk-ix
1SG.NOM see-PST-F.SG COMP there three.NOM small-GEN.PL?
- jaj'tsa liʒ-itj
egg.GEN.SG lay.PRS-3SG
- ‘I saw that there were three small eggs lying there’.

¹²⁴ According to Greville G. Corbett (p.c.) SINGULAR verbal agreement may have been motivated by the combination of conditions favouring a single verb, namely, an inanimate subject and a verb of position.

In the same vein, as when we switch from lower numerals to higher numerals, there is a ‘change of gear’ when we move from PRESENT and FUTURE tense forms to the PAST forms; there is a big shift in the system. According to Sussex & Cubberley (2006: 243):

“[...] in East Slavic, the past tenses are formed from *l*-participles and do not mark [PERSON], having lost the auxiliary. The present and future do mark [PERSON], and indeed share the same inflexions for the present and future perfective [...]” (Sussex & Cubberley 2006: 243).¹²⁵

That is to say, for the forms of the PAST TENSE, Eastern Slavonic (and so WP) verbs only distinguish GENDER and NUMBER (but not PERSON) in the SINGULAR; and only NUMBER (but neither GENDER nor PERSON) in the PLURAL (see (§3.2.2.)).

Numeral phrases containing lower numerals (with ADNUMERATIVES), as well as higher numerals and collective numerals, frequently govern NEUTER SINGULAR (especially with transitive verbs) in the PAST TENSE regardless of the GENDER of the subject (e.g. in (87) is FEMININE and in (88) MASCULINE):

- (87)
- | | | | | | |
|----|--------------------------------|----------------|--------------------------|----|---------------|
| a. | odn-a | 'dʒɪʊʈʃɪna | zaxova- l-a -s(a) | ʊ | pofɪrɪb-ɪ |
| | one-NOM.SG.F | girl(F).NOM.SG | hide-PST-F.SG-REFL | in | cellar-LOC.SG |
| | 'A girl hid in the cellar.' | | | | |
| b. | dʒɪʊ'kɪ | | zaxova- l-i -s(a) | ʊ | pofɪrɪb-ɪ |
| | girl(F).NOM.PL | | hide-PST-PL-REFL | in | cellar-LOC.SG |
| | 'The girls hid in the cellar.' | | | | |
| c. | dv-i | 'dʒɪʊʈʃɪn-ɪ | zaxova- l-o -s(a) | ʊ | pofɪrɪb-ɪ |
| | two-NOM.F | girl(F).ADNM | hide-PST-N.SG-REFL | in | cellar-LOC.SG |
| | 'Two girls hid in the cellar.' | | | | |

¹²⁵ Reproduced with small typographic modifications.

- (88) a. (odin) brat robi-v za mnoju
 one.NOM.SG.M brother(M).NOM.SG work-PST.M.SG with 1SG.INS
 ‘A brother used to work with me.’
- b. dv-a 'brati robi-l-o za mnoju
 two-NOM.M brother(M).ADNM work-PST-N.SG with 1SG.INS
 ‘Two brothers used to work with me.’
- c. pjat bra't-iv robi-l-o za mnoju
 five.NOM brother(M).-GEN.PL work-PST-N.SG with 1SG.INS
 ‘Five brothers used to work with me.’
- d. njak-i bra'ti robi-l-i za mnoju
 some-NOM.PL brother(M).NOM.PL work-PST-PL with 1SG.INS
 ‘Some brothers used to work.’

Yet, I have also found examples in the corpus of natural speech (besides the ones obtained from direct elicitation) where the NumP containing a noun in ADNUMERATIVE appears with a verb in PLURAL, which deserve to be mentioned (89).

- (89) (Tor1.25 06:51)
- tri uzrozl-ix tʃolo'viki xodi-l-i do Djad-i Volod-i
 three.NOM old-GEN.PL man.ADNM go.IPFV-PST-PL to uncle-GEN.SG V-GEN.SG
 ‘Three old men used to go to Uncle Volodja’s.’

This type of verbal agreement is common to other types of NumPs in West Polesian, so the NEUTER SINGULAR can be simply analysed as the DEFAULT form. The motivation behind the choice of one marking over another is an ongoing question I am trying to solve. According to Corbett’s (1978d, 2000) predictions, the higher the numeral, the less semantically relevant is the agreement, and thus more likely to be singular. “[...] as we move up to the scale of higher numerals, plural predicate agreement becomes

less frequent and singular agreement more frequent” (Corbett 1978d: 59). This leads Corbett to propose the following universal:

“If in a given language numeral expressions allow predicate agreements of different numbers, and if the option depends at least in part on the numeral involved, then it will be the phrases involving lower numerals which take the semantically justifiable number in the predicate” (Corbett 1978d: 368).

Corbett’s prediction is well supported, particularly in the Slavonic family. It is true that plural agreement appears less frequently with higher numerals in West Polesian. Although this only explains part of the story, since both verbal agreement forms are available for NumPs containing both lower and higher numerals.

“This type of agreement is best analysed as a ‘default’ form: the neuter singular is used when agreement is required with an element which lacks the features necessary for agreement [...] This is the default form of agreement which occurs when normal agreement is impossible” (Corbett 1983: 216, 219).

Others (e.g. Kim 2009) have justified the choice of one form over another (in Russian) by specificity (i.e. if the NP is specified, it will take PLURAL agreement; whereas, if it is unspecified/undetermined DEFAULT agreement). I used to be more inclined towards this second position (for West Polesian) for a while, although I have gathered enough sentences that do not allow me to assert such a position.¹²⁶

¹²⁶ Or at least, specificity may not be the only factor determining the choice of the GENDER/NUMBER marking in the verb.

So far, the best explanation to me is that the main motivation behind the choice is related to the lexical semantics of the verb. Transitive verbs prefer DEFAULT agreement (as we have seen in (34) a.¹), whereas intransitive verbs (except for *buti* ‘to be’), particularly telic verbs like ‘to pass away’, prefer PLURAL agreement (90).

(34) a.¹ (B6.el)

dv-a	tʃolo'viki	z	Minsk-a	kupi-l-o
two-NOM.M	man.ADNM	from	Minsk-GEN.SG	buy-PST-N.SG
 xat-u	 sobi v nas.			
house-ACC.SG	REFL in 1PL.GEN			

‘Two men from Minsk bought a house in our [village].’

(90) a. (B18.3 00:01)

bu-l-a	u	materi	odn-a	dotʃk-a;	si'n-ru
be-PST-F.SG	in	mother-GEN.SG	one-NOM.SG.F	daughter-NOM.SG	son-GEN.PL
 bu-l-o	 pjat,	 a	 dotʃk-a	 vsjofo	 odn-a.
be-PST-N.SG	five.NOM	but	daughter-NOM.SG	just	one-NOM.SG.F

‘[The] mother had one daughter. She had five sons, but only one daughter.’

b. (TL6.1 05:38)

tam	tʃolo'vik	musit	bolʃe	jek	ʃesdisjat(j)
there	person.GRADNM	maybe	more	than	sixty
 v	 t-ij	 ʃkol-i	 bu-l-o		
in	that-LOC.SG	school-LOC.SG	be-PST-N.SG		

‘There were possibly more than sixty people in that school.’

c. (B6.el)

dv-a **'xloptsɪ** **zafɪnu-l-ɪ** ʊ avar-ɪj (*zafɪnu-l-o)
two-NOM.M boy.ADNM die.PRF-PST-PL in accident-LOC.SG die.PRF-PST-N.SG

‘Two boys passed away in the accident.’

d. (B6.el)

pjetj **bab** **pomer-l-ɪ** z xolod-u (#pomer-l-o)
five.NOM granny.GEN.PL die.PRF-PST-PL from cold-GEN.SG die.PRF-PST-N.SG

‘Five grannies died from cold.’

Finally, Madariaga & Igartua (2017) explain that both NEUTER SINGULAR and PLURAL agreements are also possibilities in Russian, but that there are certain conditions that favour the use of PLURAL:

“Plural agreement is more likely to be triggered when one or more of the following factors concur: the subject precedes the verb; the subject is specific; the elements included in the subject are individuated; the numeral is low; the subject is animate; the verb is active; or the subject is feminine (Graudina et al. 1976, Kuz’minova 2004, Švedova 1980, 242– 43)” (Madariaga & Igartua 2017: 102) [See references there].

Nonetheless, I recognise the limitations of my approach and I do not want to frame such a hypothesis in a way that it excludes other potential factors conditioning the choice. So as Madariaga & Igartua (2017) suggest, it may be a combination of factors.

5.3. Nouns headed by higher numerals

In this section I focus on nouns and NPs headed by higher numerals trying to answer this question: is there a GREATER ADNUMERATIVE form? I start by looking at evidence in texts and speakers' intuitions that triggered the question, which point to the existence of a morphophonologically dedicated GREATER ADNUMERATIVE form (§5.3.1.). Then, I present the properties of this potential adnumerative form by contrasting it with the (LOWER) ADNUMERATIVE form (§5.3.2.). I then move to the opposite position and present arguments against the existence of a morphophonologically dedicated GREATER ADNUMERATIVE form (§5.3.3.). I end up this section with the conclusions extracted from the analysis, which I already advance that reject the idea that the special forms observed are a full or dedicated GREATER ADNUMERATIVE form (§5.3.4.).

5.3.1. Introduction

In the previous sections, I have shown that numeral phrases in West Polesian trigger complex agreement patterns (particularly those containing lower cardinal numerals). If nouns in NumPs with lower numerals took ADNUMERATIVE, nouns in NumPs with higher numerals seem to also have a dedicated form available: the GREATER ADNUMERATIVE (hereafter, GRADNM). For example, in certain contexts, such as (91) a-b. (with approximate numerals), the use of the regular GEN PL (which in these examples is suppletive) is not permitted:¹²⁷

¹²⁷ More on this in (§4.2.3.).

- (91) a. (B20.8 00:11) [...]
 zbira-l-a-sa kompanja; **tsolo^lvik** **deset** [...]
 gather-PST-F.SG-REFL company-NOM.SG person.GRADNM ten.NOM
 (*lud-ej deset)
 person-GEN.PL ten.NOM
 '[...] people used to gather; about ten people [...].'
- b. (TL1.1 16:45)
duš **djesjet** na^lvjerno [...] (*ludej deset)
 person.GRADNM ten.NOM probably person-GEN.PL ten.NOM
 'About ten people, probably, [...].'

However, most of the time, the potential GREATER ADNUMERATIVE (GRADNM) form is, at least, homophonous with the GENITIVE PLURAL (GEN.PL) form (92) a-b.:

- (92) **Answers given by B13-B14 to visual stimuli and direct elicitation**
- a. *pjet* **'ptfolej** [GRADNM?] 'five bees'
- b. *numa* **'ptfolej** [GEN PL] 'there are no bees'.

For this reason, the nature of this form is far less obvious, and thus, more arguable, than the ADNUMERATIVE. In fact, I only started becoming aware of it in the last stages of my research.¹²⁸

5.3.1.1. Evidence from Elicitation

Most of the forms suspected to have a GREATER ADNUMERATIVE have been extracted from free texts, contrasted with a considerable amount of direct and semi-direct

¹²⁸ Time and visa limitations prevented me from devoting as much attention to this phenomenon as I would have liked. Thus, all the conclusions for this parameter are highly provisional.

elicitation (i.e. visual prompts). The speakers were presented with different numbers of animals and they were asked to give the numeral with the corresponding form of the noun (as it was partly conducted for the LOWER ADNUMERATIVE, initially). Surprisingly, very often speakers proposed different forms for what I considered to be a single cell (GEN PL) (see (91) *supra*).

I asked several speakers from Bahdanaŭka about their choice.¹²⁹ I obtained different responses about the underlying rule, which made me suspect for a long time that there was an emergent internal division within the group of higher numerals, but whose boundaries were relative (speakers' perception of smaller or bigger quantities). Therefore, this would contrast with the (LOWER) ADNUMERATIVE where the boundary for 'two', 'three' and 'four' is well defined.

I still struggle to find any consistency or evidence to state that such a division is real, although I would not like to disregard the speakers' intuitions about a special form. So far, I can only say that I have observed a remarkable tendency in B6 & B9 (mother and son) to use zero endings (-∅) or -iv (for inflectional class II nouns) when the noun is headed by a higher numeral (but not very high), and the form -ej as the elsewhere form. It may be a similar correlation between individuation and the form chosen as it has been suggested for verbal agreement; i.e. the higher the quantity, the less likely for the entities to be distinguished, more "nouny" and more of a "straightforward GENITIVE PLURAL" (Corbett 1978b, 1983). And the fact that different speakers have

¹²⁹ Bahdanaŭka is where this phenomenon was first documented, and where it seems to be the most prominent among the villages surveyed.

shown an intuition about the preferences of forms should make us suspect the existence of something more than mere overabundance.

5.3.1.2. Evidence from Non-Canonical Phenomena

Cells closely related to numerals and quantification seem to behave like a Bermuda Triangle for non-canonical inflectional phenomena. In all the data analysed it seems that non-canonical phenomena are especially prone to occur in adnumerative forms. In particular, suppletion and heteroclisis, the most prominent, warn us that this may not be a straightforward GEN PL form, and that there may be a deeper phenomenon underneath these cell(s). For example as I show in the next chapter (Chapter 6) the nouns ‘year’ and ‘person’ often have ‘remainders’ (using Corbett’s (2007) terminology) in what looks like the GEN PL cell. The noun for ‘person’ has the [NOM SG] *tʃolo'vɪk*, and for the majority of the speakers I have interviewed, a suppletive [GEN PL] *lu'dej*, but a heteroclitic [GRADNM] *tʃolo'vɪk*. Yet, if the GRADNM and the GEN PL forms were purely overabundant (and redundant), and thus, fully interchangeable, then there would not be contexts excluding the use of one of the forms. See (93):

- (93)
- | | | | |
|------------------|-----------------------|------------------|------------------|
| a. ^I | pjetj | lu'd-ej | |
| | five.NOM.PL | person-GEN.PL | |
| a. ^{II} | pjetj | tʃolo'vɪk | |
| | five.NOM.PL | person.GRADNM | |
| | ‘Five people.’ | | |
| b. ^I | saraj | fetix | lu'd-ej |
| | barn.NOM.SG | this.GEN.PL | person-GEN.PL |
| b. ^{II} | * saraj | fetix | tʃolo'vɪk |
| | barn.NOM.SG | this.GEN.PL | person.GRADNM |
| | ‘These people’s barn’ | | |

In examples (93) a.¹-a.^{II} both forms (involving suppletion) are synonyms; but we see in (93) b.^{II} that [GRADNM] *tfolovik* cannot be used as a straightforward GENITIVE PLURAL form (i.e. on its primary sense of denoting belonging to someone).

Moreover, the use of the forms *tfolovik* ‘person’ or *hod* ‘year’ for GRADNM makes them heteroclitic (more on this in (Chapter 6)). Since they belong to infl. class II, we expect them to take either *-iv* or *-ej* (like other GEN PL forms), and not zero marking, which is proper of infl. class I and III nouns. Far from being a coincidence, this seemingly morphomic pattern also repeats in itself with the alternative suppletive stem of ‘year’ *rk-*. So, it is certainly curious that in all the three cases heterocclisis makes the GRADNM syncretic or at least homophonous with the NOM SG form.

Thus, there are two possible hypotheses or analyses for this form:

H5: There is a GREATER ADNUMERATIVE, which has a phonologically distinct form for certain nouns.

H6: All the forms mentioned are GEN PL (admitting a high level of overabundance), although under certain conditions there is a remarkable tendency for specific forms to be used (particularly, the ones that are not the predictable GEN PL).

H5 is more in line with the broad definition of ADNUMERATIVE as stated at the beginning. Yet, as noted earlier, for the sake of clarity, I will only call ADNUMERATIVE those forms of the paradigm that are morphophonologically dedicated. However, before going into the description, I need to make a small terminological clarification.

5.3.1.3. Terminological Remarks: why it is misleading to use the terms ADNUMERATIVE SINGULAR and ADNUMERATIVE PLURAL?

According to Mel'čuk (1985: 430-437) a handful of nouns in Russian have an ADNUMERATIVE PLURAL form. Mel'čuk argues that the ADNUMERATIVE PLURAL forms appear when the nouns are headed by a higher numeral (or quantifier), although they can be replaced by the predictable GEN PL. For example, when the noun *kilogram* appears headed by a higher numeral both the predictable GEN PL (-ov) and ADNUMERATIVE PLURAL are allowed: [5] *kilogramm*ov / *kilogramm*. Nevertheless, it is important to highlight that except for two nouns on that concise wordlist, the rest are technical words such as *gauss* or *rentgen* (Roentgen) which entered the language roughly after 1900.¹³⁰ These nouns had unusual codas (which were hard to process for Russian phonotactics), which seems to be the origin of the confusion over the inflection and the overabundance of forms (the GEN PL/ ADNM PL of these forms is -∅ (zero) or lack of inflection).¹³¹ Moreover, Mel'čuk (1985) points out that nouns having an “adnumerative plural” do not have an “adnumerative singular”, and vice versa, which is already a bit suspicious.

In Russian it may seem a coherent or consistent form to denote the analogous form ADNUMERATIVE PLURAL as Mel'čuk (1985) did. The Russian (LOWER) ADNUMERATIVE form is almost always syncretic with the GENITIVE SINGULAR form; in comparison with the alleged GREATER ADNUMERATIVE form which for most nouns is syncretic with the

¹³⁰ The only exceptions are the form *čelovek* (cognate of WP *tfolovik*) ‘person’ (more on this noun in Chapter 6)), and *cvetok* flower.

¹³¹ Matthew Baerman (p.c.) has suggested that the origin of this form could be in semantic generalisation, i.e. things that get counted.

GENITIVE PLURAL. That is to say, the similarity with well attested forms is probably the motivation behind that **NUMBER** distinction.

Now, when it comes to West Polesian, the (LOWER) **ADNUMERATIVE** is a knotty form as it displays properties of **CASE** and **NUMBER** values. Thus, the first objection for calling this form ‘**ADNUMERATIVE PLURAL**’ is that it necessarily implies that the **ADNUMERATIVE** is a **CASE** value, and that this form is its **PLURAL** counterpart (I discuss this further in (§5.5.2.)). Related to this, my second objection is related to the fact that calling the (LOWER) **ADNUMERATIVE** ‘singular’ is inaccurate. In the previous section (§ 5.2.3.) I have provided evidence that NumPs with lower numerals, which contain (LOWER) **ADNUMERATIVE** forms behave like **NON-SINGULAR** in terms of morphosyntactic agreement. And furthermore, the **ADNUMERATIVE** itself derives etymologically from a **DUAL** (and thus, retains part of the **NON-SINGULAR** semantics). In addition, even though morphophonological patterns suggest a link with the **SINGULAR** subparadigm, I have shown how these patterns can be inconsistent and misleading in West Polesian, and particularly the adnumerative cells being on the fringe of the paradigm are more akin to deviations. Consequently, calling this (greater) adnumerative form **PLURAL** without a clear **SINGULAR** creates a gap in the proposed system.

Finally, admitting this a weaker argument, if we believe that the form that collective numerals are taking is the same as the one we are calling **GREATER ADNUMERATIVE**, then it is problematic to explain how an adnumerative plural can also cover the domain of lower numerals (semantically). For example, many speakers in Bahdanaŭka produced pairs like (94) a-b. as synonyms:

(94) **Pair suggested by speakers in Bahdanaŭka**

- | | | | | | |
|----|------------------|-------------|----|-------------|-----------------------|
| a. | tri | 'doktarɪ | b. | troje | doxta'r-iv |
| | three.NOM | doctor.ADNM | | three(COLL) | doctor-GEN.PL/GRADNM? |
| | 'Three doctors'. | | | | |

In sum, for these reasons, I have decided to borrow the terms LOWER and GREATER from Corbett's (2000) description of different types of PAUCALS. Such a terminology implies a relation between the two forms, but it does not bind them tightly to a particular analysis, while also allowing them to manifest differently across paradigms. Note that for practical reasons (among which, the fact that there is solid evidence that proves the existence of the (LOWER) ADNUMERATIVE, whereas this is not the case for the GREATER ADNUMERATIVE) I refer to the LOWER ADNUMERATIVE simply as ADNUMERATIVE.

5.3.2. Properties of the GREATER ADNUMERATIVE

5.3.2.1. Similarities with the ADNUMERATIVE

In spite of the previous point, there are significant similarities between the ADNUMERATIVE and the alleged GREATER ADNUMERATIVE, which may be an indicator of a correlation between both forms. And again, any ultimate solution about the status of the adnumerative (particularly the lower adnumerative) cannot take this for granted.

5.3.2.1.1. Availability

The dedicated GREATER ADNUMERATIVE under discussion is only available for nouns, as it also applies to the (LOWER) ADNUMERATIVE.

5.3.2.1.2. Syntactic and semantic autonomy

Both forms are related to quantification and are used with numerals. Nevertheless, none of them are either semantically or syntactically fully autonomous (95) a-b.:

- (95) a. pjet **traxto'r-ej/iv** volotʃɪ-l-o polj-e
 five.NOM tractor.GRADNM/GEN.PL plow-PST-N.SG field-ACC.SG
 ‘Five tractors ploughed the field’.
- b. * **traxto'r-ej/iv** volotʃɪ-l-o polj-e
 tractor.GRADNM/GEN.PL plow-PST-N.SG field-ACC.SG
 ‘(A certain amount of) tractors ploughed the field’.

5.3.2.1.3. Verbal agreement

Numeral phrases headed by higher and lower numerals (and thus, containing nouns potentially in adnumerative forms) follow an almost identical verbal agreement pattern, which is also in line with the overall behaviour of other numeral phrases and numerals (collectives and pronominals) in West Polesian.

In the PRESENT TENSE NumPs containing higher numerals (and thus, potentially a GRADNM) govern 3RD PL agreement:

- (96) pjetj **xlopts-uv** rob-ljatj
 five.NOM boy-GRADNM/GEN.PL work-3PL
 ‘Five boys are working’

Nevertheless, in the FUTURE TENSE, particularly when the subject appears after the verb, the verb usually takes 3RD SG:

(97) (Z4.5 03:30)

a	tiper	ni-ma	ko'riv;	ka3utj	na [...]	vs-e
and	now	NEG-HAVE	cow.GRADNM/GEN.PL	say-3PL	in	all-LOC.SG?
3ɪdtʃ-e	bud-e	vsjofo	devet	ko'riv		
Ž-LOC?	be.FUT-3SG	just	nine	cow.GRADNM/GEN.PL		

‘There are no cows anymore. They say there are only nine cows in Žydča
(lit. ‘there will only be only nine cows’).

This is a common phenomenon with phrases involving equations, with parallels in other Slavonic languages (at least in Eastern Slavonic) (Pugh & Press 1999, Shevelov 1963).

In the PAST TENSE, both default NEUTER SG and PLURAL agreement are possible, regardless of the GENDER of the head, as happened with lower numerals. Otherwise, I will not dedicate much time explaining verbal agreement:

(98) (B6.el)

pjetj	3anotʃin	kupɪ-l-o	fiazet-u	[kupɪ-l-ɪ]
five.NOM	woman(F)-GEN.PL	buy-PST-N.SG	newspaper-ACC.SG	buy-PST-PL

‘Five women bought the newspaper.’

In sum, the NumPs containing higher cardinal numerals (and hence, containing a noun in the potential GREATER ADNUMERATIVE) follow the same verbal agreement (and a similar but more limited nominal agreement) as those containing lower numerals (and thus, a noun in ADNUMERATIVE). Yet, according to the cross-Slavonic survey carried out by Corbett (1983: 220-224), PLURAL predicate agreement decreases “monotonically” the higher the numeral in *all* Slavonic languages in the survey.¹³²

¹³² Corbett (1983: 222) remarks that the tendency is less remarkable in certain groups, in particular in the Eastern Slavonic group where “the situation is more fluid”.

His claims are based on statistics, elaborated on a corpus research of text from different Slavonic languages. Thus, given that I do not have a corpus big enough to replicate the experiment with West Polesian, we will assume that what applies to all other Slavonic languages is probably true for West Polesian as well. Having said this, the tendency to trigger SINGULAR predicate agreement is a continuum which also affects lower numerals. So, in this respect, phrases containing ADNUMERATIVES and GREATER ADNUMERATIVES behave the same way.

- (99) a. pjet xoroʃ-ix bra't-iv
 five.NOM beautiful-GEN.PL brother(M)-GEN.PL
 'Five handsome brothers.'
- b.ⁱ pjetj xoroʃ-ix di'vok
 five.NOM beautiful-GEN.PL girl(F).GEN.PL
- b.ⁱⁱ *pjet xoroʃ-ie di'vok
 five.NOM beautiful-NOM.PL girl(F).GEN.PL
- b.ⁱⁱⁱ *pjet xoroʃ-ie diu'ki
 five.NOM beautiful-NOM.PL girl(F).NOM.PL
- 'Five beautiful girls.'

5.3.2.1.4. NP Agreement

Both forms allow adjectives (and other elements in the NP) to stand in (at least, what resembles) GENITIVE PLURAL, regardless of the GENDER of the noun they specify. The (LOWER) ADNUMERATIVE has the possibility of using the NOMINATIVE PLURAL (with an effect on the noun as well), whereas this is not permitted with the GREATER ADNUMERATIVE (or GEN PL).

5.3.2.2. Differences with the ADNUMERATIVE

As I have previously stated before, even if I can prove the GREATER ADNUMERATIVE is a real form (i.e. a morphosyntactic value with a morphophonologically dedicated form), there are still significant differences between the ADNUMERATIVE and the GREATER ADNUMERATIVE.

5.3.2.2.1. Range of Use

The first most striking difference between both adnumerative forms is that in theory the alleged GREATER ADNUMERATIVE is also available for quantifiers other than numerals, such as *how many*?

Forms proposed by B2 and also B6 for *korova* (F) ‘cow’¹³³

- | | | | | | | |
|-------|----|----------------------|------------------|----|----------------------------|---------------------------|
| (100) | a. | <i>pjetj ko'rɪv</i> | ‘five cows’ | d. | <i>mnifā koro'vej</i> | ‘many cows’ |
| | b. | <i>desɪtj ko'rɪv</i> | ‘ten cows’ | e. | <i>skɪ(l)ke ko'rɪv</i> | ‘how many cows’ |
| | c. | <i>sto ko'rɪv</i> | ‘a hundred cows’ | f. | <i>moloko tɪx koro'vej</i> | ‘milk of those cows’ [B6] |

However, this has not been the case for other quantifiers, such as ‘many’ or ‘much’, as speakers proposed the regular GEN PL after these, as we can see from (101). For example, I have heard different speakers saying *mnifā fiadov* ‘many years’ instead of *mnifā fiad* which is the form they use with numerals.

In this respect, the nature of this adnumerative is more inclusive with governors, but the term adnumerative is less appropriate, as this form is not exclusive to NUMERACY

¹³³ Note that although B2 and B6 are from the same village, there was a gap between B2’s only interview and the interview in which I elicited these data from B6.

stricto sensu. In this respect, West Polesian GREATER ADNUMERATIVE typologically approaches the adnumerative form that Bulgarian, Macedonian and Ossetic (Sims-Williams (forthcoming)) have (see Table 18).

Table 18 Scope of the adnumerative forms in WP and Bulgarian/Macedonian

	LOWER NUMERALS	HIGHER NUMERALS	QUANTIFIERS
West Polesian	ADNUMERATIVE	GREATER ADNUMERATIVE	
Bulgarian/Macedonian	ADNUMERATIVE FORM		

5.3.2.2.2. Etymology

The West Polesian ADNUMERATIVE is begotten of an eroded NOM/ACC DUAL, and retains properties of the CASE form as well as the NUMBER form (and there is diachronic evidence for this). By contrast, the origin of the GREATER ADNUMERATIVE is more controversial and thus it can be used as an argument against its existence. On the one hand, there is enough evidence that higher numerals used to take what looks like GEN PL in Common Slavonic (Mel'čuk 1985, Suprun 1961, Žolobov 2003), and there seems to be a wider consensus for this in the literature. On the other hand, the emergence of a new form for these syntactic contexts (other than a straightforward GEN PL) may have been motivated by analogy with the LOWER ADNUMERATIVE, making use of the 'spare' or overabundant forms of the GEN PL of nouns. Nevertheless, the GREATER ADNUMERATIVE can only be an emerging category in morphology.¹³⁴

¹³⁴ Because it would still be a lot less morphologically robust, and thus regular, than the ADNUMERATIVE.

5.3.2.2.3. Morphological Robustness

As I have previously mentioned, the GREATER ADNUMERATIVE is only available for nouns, and not every (countable) noun has a dedicated form (as I have already mentioned in (§5.3.1.) and will develop further in (§5.3.3)), which is a non-canonical (and thus suspicious) behaviour. Conversely, we have seen that a dedicated ADNUMERATIVE form is potentially available for every countable noun (although, depending on the position of the stress it may not be phonologically distinct from other cells).

Regardless of whether we eventually prove the existence of a (dedicated) GREATER ADNUMERATIVE form or not, this form lacks a consistent form of exponence. This does not happen with the (LOWER) ADNUMERATIVE, which always (or in all the analysed cases) consists of the stem (and stress pattern) of OBLIQUE SINGULAR cells + the suffix *-i*. Thus, the alleged GREATER ADNUMERATIVE is far less canonical in this respect.

I discuss the morphological evidence more in order to disprove its existence and provide further examples in (§5.3.3.).

5.3.2.2.4. ANIMACY

Whilst the ADNUMERATIVE often shows some type of sensitivity to ANIMACY (admitting it can also override it), the GREATER ADNUMERATIVE, because of its morphology, is insensitive to ANIMACY constraints.¹³⁵ That is to say, there are no ANIMACY differences in the marking of the object with higher numerals, whereas they can appear with lower numerals (101):

¹³⁵ The GREATER ADNUMERATIVE is almost always syncretic with GEN PL, and thus also ACC PL = GEN PL.

(101) (B6&B9.el)

- | | |
|---|---|
| <p>a.^I batʃ-u dv-a 'smɪ
 see-1SG two-NOM.M son.ADNM</p> <p style="text-align: center;">[dva 'sina]
 two-NOM.M son.GEN?SG</p> <p>‘I see two sons.’</p> | <p>a.^{II} batʃ-u dv-ux sɪ'n-ɪv
 see-1SG two-ACC son-ACC.PL</p> |
| <p>b.^I *batʃ-u pjet sɪ'nɪ
 see-1SG five.NOM son.NOM.PL</p> <p>‘I see five sons.’</p> | <p>b.^{II} batʃ-u pjet sɪ'n-ɪv
 see-1SG five.NOM son-GEN.PL</p> |
| <p>c.^I batʃu dv-a 'dubɪ
 see-1SG two-NOM.M oak.ADNM</p> <p>‘I see two oak trees.’</p> | <p>c.^{II} *batʃ-u dv-ux du'b-ov
 see-1SG two-ACC oak-GEN.PL</p> |
| <p>d.^I *batʃ-u pjet 'dubɪ
 see-1SG five.NOM oak.ADNM</p> <p style="text-align: center;">[*du'bɪ]
 oak.NOM/ACC.PL</p> <p>‘I see five oak trees.’</p> | <p>d.^{II} batʃ-u pjet du'b-ov
 see-1SG five.NOM oak-GEN.PL</p> |

5.3.2.2.5. Optionality

Both the ADNUMERATIVE and the GREATER ADNUMERATIVE are quite versatile (i.e. they can be replaced by ‘regular’ GEN SG/PL). When cardinal numerals are used postnominally (as approximate numerals) the (LOWER) ADNUMERATIVE is outlawed and the GREATER ADNUMERATIVE is the only form permitted (even if the numeral is a lower one), but not the GEN PL, as I have shown in (91) (*supra*).

5.3.2.3. Summary of Comparisons with the ADNUMERATIVE

I present a summary of the differences and similarities in Table 19:

Table 19 Differences between the ADNUMERATIVE and the GREATER ADNUMERATIVE

	ADNUMERATIVE	GREATER ADNUMERATIVE
Availability for parts of speech	Only for nouns	Only for nouns
Semantic and syntactic autonomy	Non-autonomous	Non-autonomous
Verbal NUMBER agreement (PRESENT TENSE) of such NumPs	PLURAL	PLURAL (but, increasingly SINGULAR the higher the numeral)
Preferred verbal NUMBER/GENDER agreement (PAST TENSE) of such NumPs	DEFAULT	DEFAULT
Range of use	Lower (cardinal) numerals	Higher (cardinal) numerals, collectives and (some) quantifiers
Etymology	Eroded NOM/ACC DUAL	Uncertain, GEN PL and/or analogy with the adnumeration
Exponence	Regular suffix attached to OBLIQUE SINGULAR stem	Heterogeneous
Sensitivity to ANIMACY	Partly	No
Optionality	Yes	No (in certain contexts)

In comparison to the (LOWER) ADNUMERATIVE, this form is far more regular in terms of syntactic agreement, although morphologically it is much less robust.

5.3.3. Arguments against the existence of the GREATER ADNUMERATIVE

In order to test whether there are any tendencies for the different forms in the GEN PL/GRADNM cell(s) I have set different syntactic/semantic conditions.

5.3.3.1. Possession

Let us start with the straightforward GEN PL form; i.e. when it is used for its primary function ‘to denote belonging or relation to something or someone’. The forms in Table 20 were elicited from different speakers of the same village (Bahdanaŭka) within a range of a few weeks. The prompts for the forms were ‘The honey of those bees’, for the GEN PL, and images of five and ten bees.

Table 20 Results from survey on the noun *ptŋola*(F) ‘bee’ in Bahdanaŭka

Speaker	GEN POSSESSIVE	HIGHER NUMERAL	NOM PL
B6	'ptŋolej	[5] ptŋıl [10] 'ptŋolej/ ptŋıl	'ptŋolı
B10	[n.d.]	[5/10] 'ptŋolej	'ptŋolı
B11 (&B12)	'ptŋolej	[10] 'ptŋolej	'ptŋolı
B13 (&B14)	[n.d.]	[10] 'ptŋolej	'ptŋolı
B15	'ptŋoliu	[10] 'ptŋolej	'ptŋolı
B17	'ptŋolej	[6] ptŋo'lej [10] ptŋol [10/20] 'ptŋolej ¹³⁶	'ptŋolı

In Table 20 we can observe an absolute consensus on the NOM PL form. Regardless of the fact that there was a general preference for one GEN POSSESSIVE form, the results show how much variation is possible particularly with numerals. The forms *ptŋıl* and *ptŋo'lej*

¹³⁶ Only after B9's suggestion.

seem exclusive for numerals, but it is true that they only appear isolated in two idiolects. The form *ˈptʃoliu* for GEN POSSESSIVE seems also restricted to this function, but since it has only been documented in one idiolect there is not enough data to contrast it.

5.3.3.2. Polarity

Another common use of the GENITIVE in Slavonic languages is to use it with negation. West Polesian uses the GENITIVE OF NEGATION less than other Slavonic varieties (e.g. Polish). But in constructions with *nma* (in any tense), meaning ‘there is not’, the use of the GENITIVE is obligatory. When compared to the forms appearing with higher numerals, the results are shown in Table 21.

Table 21 Variations for the noun ‘cow (F)’

Speaker	LOWER NUMERALS	HIGHER NUMERALS	NEGATION	NOM PL
B6	ko'rovi	koro'vej/ ko'riʋ	koro'vej	koro'vi
B9	ko'rovi?	ko'riʋ	koro'vej	koro'vi
B11	[n.d.]	koro'vej	ko'rovej ¹³⁷	koro'vi
B12	ko'rovi	ko'rou	ko'rou	[n.d.]

The sample is very small and speakers are related to each other (B6 to B9, and B11 to B12), which may explain the homogeneity. It is surprising that none of the participants interviewed used the form *ko'riʋ* on a negative sentence, although I have found one speaker from the same village using *vou'kiʋ* with a negative (which was

¹³⁷ B11 also started saying *koro'vej* after hearing it from B9.

directly elicited). Again, we see homogeneity in the NOM PL cell and even the ADNOM cell (which is less common), but when it comes to the cells around GEN PL, they display a considerable amount of inconsistency and variation (see (98) *supra* and (103) *infra*).

(102) (Z.4.2 08:25)

tam zare musit i ni-ma i pɪtdɪsʲatʲ utʃ'n-iv
 there now maybe and NEG-HAVE and fifty.NOM student-GEN.PL
 'Probably now there are less than fifty students.'

5.3.3.3. Quantifiers

The use of the GENITIVE as a PARTITIVE seems a common phenomenon in European languages (e.g. Greek). Certainly most Slavonic languages use GENITIVE PLURAL when the noun is preceded by a quantifier such as 'many' or 'little'. In Table 22 and Table 23 I have summarised different responses given by the speakers in Bahdanaŭka comparing the contexts where they use a quantifier (with the nouns 'dog' and 'chicken' as a representative) with those where they use a higher numeral:

Table 22 The noun 'dog' in different contexts

Speaker	LOWER NUMERALS	HIGHER NUMERALS	QUANTIFIERS	NOM PL
B6	so'baki	[n.d.]	so'bakej /soba'kej /so'bak	soba'ki
B10	so'baki	so'bakej	so'bakej	[n.d.]
B11	so'baki	so'bakej	so'bakiu ¹³⁸	so'baki
B13	so'baki	[n.d.]	so'bak	so'baki
B17	so'baki	so'bak	[n.d.]	[n.d.]

¹³⁸ Used in a context where it would be quantifier [ACC PL] + noun [ACC PL]

Table 23 The noun ‘chicken’ in different contexts

Speaker	LOWER NUMERALS	HIGHER NUMERALS	QUANTIFIERS	NOM PL
B6	'kuri	[5, 6] kur [10] ku'rej	ku'rej	'kuri
B10	'kuri	ku'rej	ku'rej	[n.d.]
B11 &12	'kuri	ku'rej (B11); kur (B12) ¹³⁹	[n.d.]	'kuri
B13	'kuri	ku'rej	ku'rej	[n.d.]
B15	'kuri	ku'rej	ku'rej	'kuri
B17	'kuri	kur	kur ¹⁴⁰	[n.d.]

The data in Table 22 and Table 23 are less complete than for others. Table 23 displays a significant consistency of forms. Yet, for Table 22 it is hard to extract any meaningful conclusions, especially because there is not even homogeneity in the NOM PL cell (there is absolute homogeneity on the ADNUM cell though), which should alert us slightly. Except for the form in *-iv* in Table 22 identified again with quantifiers, but not numerals, there seems to be little evidence to state that higher numerals take a form that is different from the one with quantifiers.

5.3.4. Conclusions

I have not found solid morphological evidence to believe that there is a dedicated GREATER ADNUMERATIVE form. It is tempting to try to give an account of what is happening with higher numerals and based on what we know about the ADNUMERATIVE and lower

¹³⁹ After listening to B11, he started saying *ku'rej*.

¹⁴⁰ Although she used the form *ku'rej* in an ACC PL = GEN PL context.

numerals.¹⁴¹ Yet the implications of such a move overcomplicate the system, by forcing us to create multiple ‘ghost-cells’ which are very regularly syncretic with the GEN PL. Furthermore, according to Corbett (1983), the gap between lower and higher numerals used to be far more pronounced in the past (looking at the morphophonological form of the numerals), so there is no need to try to hide that gap. It seems that certain nouns have a preference for certain forms in contexts where they are quantified, which in some cases may have become lexicalised. Yet, for most nouns there is not enough solid evidence to claim for a morphosyntactically different form. So, in the light of the data, Matthew Baerman (p.c.) suggests that this form is better analysed as a sub-variety of GEN PL, like the SECOND LOCATIVE in Russian (see Brown’s (2007) analysis for more details on this). Brown (2007) says that:

“[...] the second locative is not opposed to the first locative [...] but is instead a specialization and, as such, is not in direct paradigmatic opposition with the other cases. [...] It could be true that all nouns had a separate second locative form, but these second locative would still remain a sub-case, because its appearance is limited to one number value.” (Brown 2007: 69)

All this can be applied to the GREATER ADNUMERATIVE (i.e. only available for one NUMBER value, and just for a handful of nouns). Conversely, in (§5.5.2.) I give more details why it is not correct to call the (LOWER) ADNUMERATIVE a CASE value (i.e. it would be a very non-canonical CASE value; with functions very close to quantification and derived from a NUMBER value, among other arguments).

¹⁴¹ For example, Rappaport’s (2002, 2003) approach would be in essence close to this idea. In the next section (§5.4.), I introduce different analyses that have been proposed for analogous constructions in Russian.

There is not a single exponent of the GREATER ADNUMERATIVE, whereas the dedicated (LOWER) ADNUMERATIVE is always marked by *-i* (combined with the stress for disambiguation). I admit that this last point is a much weaker argument, inasmuch as the GEN PL marking varies considerably from one inflectional class to another, and often within the same inflectional class there are various options for marking it. Yet, nobody doubts the existence of the GEN PL.

Coming back to overabundance, overabundance of GEN PL forms has already been reported for BLM and ULM (Jankoŭski 1980, Ljapěškin 1989, 1990, 1991a, 1991b, Narkevič 1976). Consequently, given that this is an ongoing process in standardised languages, we should not be surprised to find even greater variation in a non-standard form.¹⁴²

In sum, the data analysed so far show that the GREATER ADNUMERATIVE is just a SECOND GENITIVE PLURAL form that certain nouns have, and which in some specific contexts is lexicalised. There is not enough morphological or syntactic evidence to state that it is a morphophonologically dedicated form, as happens with the ADNUMERATIVE. Having said all this, I admit that this is a complex question, which, however, has not been a priority in my research. Hence, this phenomenon deserves more profound analysis to be carried out in future research.

¹⁴² For example, Narkevič (1976) made the following comment on Standard Belarusian (BLM) “The feminine [meaning infl. class I] noun *kury* [‘chicken’] takes the [GEN PL] ending *-ej*, as it has a collective meaning and it can be employed for both genders. [Note: Under the influence of the spoken language [Translators’ note: probably meaning ‘dialects of Belarusian’], often in literary works the noun *kury* appears as *kur* in genitive, with zero ending]” (Narkevič 1976: 170-171) [My translation].

5.4. Syntactic analysis: perspectives

As we have seen in the examples from Russian in the Introduction (66), NPs headed by numerals present some descriptive challenges, especially when adjectives are inserted. Probably the big underlying question is related to government and agreement: which constituent is governing or agreeing with which? Russian numeral phrases, particularly those containing a lower numeral, have been analysed and discussed by many authors working in different frameworks. However, for the past twenty years, predominantly generative linguists have tried to give a solution to the riddle. Given some of the similarities or partial overlaps with West Polesian NumPs, I have decided to give a short overview of what other colleagues have contributed to this ongoing debate.

On the one hand, Corbett (1978c, 1983), Mel'čuk (1985) and Zaliznjak (1973, 2002) have dealt with the noun form appearing after lower numerals as a GENITIVE SINGULAR with some special properties.

Corbett (1978c, 1978d) proposed that the nouns are the real heads or controllers in NumPs and he named the form that nouns take after numerals as GENITIVE. For him, the fact that nouns preceded by a numeral can take GENITIVE is an indicator of “nouniness” (of the numeral), in comparison to the use of NOMINATIVE (like ‘one’) which for him is a more “adjectival” behaviour. Thus,

“[t]he node dominating the numeral must be marked for the degree of nouniness [...] and it is this which will trigger or fail to trigger genitive insertion. Numerals like *million* always require genitive insertion while those like *pjat'* require it only to avoid double nominatives” (Corbett 1978d: 361).

Babby's (1987) analysis is based on higher numerals. Nevertheless, he answers the apparent CASE/NUMBER mismatches on numeral phrases by saying that

“while the head noun does, in fact, control the number and gender agreement of its modifiers, it does not control their case marking [...] The quantified noun [...] is the head of the NP in all cases, and the number is a modifier” (Babby 1987: 91- 102).

Pesetsky's (2013) approach relies heavily on the syntax at the expense of morphology.¹⁴³ In brief, he claims that all Russian nouns are somehow “born genitive” (2013: 9), thus the surface form of the nouns headed by lower numerals; and for this same reason nouns (can) assign GENITIVE to other constituents (Pesetsky 2013: 81).

According to Kim (2009: 22) Rappaport's (2002, 2003) approach also gives preference to syntax over morphology to solve the puzzle. Rappaport's approach (reconsidered by Kim 2009) advocates for a special (purely syntactic) case form called QUANTITATIVE (for NumPs headed by higher numerals).

On the other hand, certain authors have given more importance to the morphology, while analysing the forms appearing with lower numerals as a NUMBER form. According to Madariaga & Igartua (2017: 103) Dingwall (1969) was the first to

¹⁴³ “Because the distribution of case forms across syntactic configurations still presents numerous complex and unsolved puzzles, it is at least conceivable that the solution to one or more open questions about the syntax of case might allow us to “eliminate the middleman” after all, by reducing the case categories to independently attested properties of the syntax” (Pesetsky 2013: 6).

propose “that the seemingly genitive case associated with paucal numerals was in fact the surface manifestation of an underlying ‘dual’ marking [...]”.

Bailyn & Nevins (2008: 266) reconsidered this idea and also proposed that what on the surface looks like GEN SG is NOM PAUCAL morphology, with the addition that this “paucal morphology” is also available for adjectives, whilst verbs only distinguish SG from NON-SG in the morphology.¹⁴⁴ Their main argument is that:

“[n]ominative case and verbal agreement are biconditional in Russian. To allow plural agreement to be triggered by a genitive singular head noun would thus run counter to one of the language’s strongest exceptionless morphosyntactic generalizations” (Bailyn & Nevins 2008: 266).

Madariaga & Igartua (2017) go even further on treating the seeming GEN SG forms as PAUCALS,¹⁴⁵ based on evidence from diachrony. They show how the form that nouns take after a lower numeral has only been syncretic with (most) GEN SG forms for the last couple of centuries. What is more, the fact that Russian retains an ADNUMERATIVE form in a handful of nouns is an indicator that the underlying form is a NOM PAUCAL, for which there is a special morphology.

¹⁴⁴ They explain that from the point of view of morphosyntactic feature values, there are three values: the PAUCAL and the PLURAL, which are both [-SG], although the PLURAL has an added [+ AUGMENTED] value which is distinguished from the former one in verbal marking (this will be also adopted by Pereltsvaig (2013)). Bailyn & Nevins (2008) also remark that in BCMS (which also has a robust ADNUMERATIVE form for nouns, as we have seen in (§5.1.2.3.) the PAUCAL triggers special PAUCAL verbal agreement (-a) for MASCULINES in BCMS in the PAST TENSE.

¹⁴⁵ Although they do not share with Bailyn & Nevins (2008) that the PAUCAL value is available for other parts of speech.

“[...] due to the defectiveness of paucal number morphology in Russian, nouns use apparent genitive case forms to express it. These forms can be interpreted as exponents of non-autonomous number values [...]” (Madariaga & Igartua 2017: 104).

Many of the analyses of Russian numeral phrases have disregarded the adnumerative forms.¹⁴⁶ This can be explained by the fact that they are very marginal in Russian. Nevertheless, since adnumerative forms are morphophonologically robust in West Polesian and trigger different effects on verbal and nominal agreement, a lot of what has been said about Russian cannot be entirely applied to West Polesian.

Let us reconsider five possible configurations (103). All these phrases are possible with lower (103) a-c. and higher (103) d. numerals, yet only b. and d. contain proper adnumerative forms, following the definition and criteria set at the beginning. Note that as noted in the previous section (§ 5.3.4.), most likely proper GREATER ADNUMERATIVE forms do not exist. In any case, I mark the noun in (103) d. as having two possible interpretations (i.e. CASE/NUMBER values). In addition, (103) a. and (103) b. are homophonous when inflection class I nouns are inserted.

¹⁴⁶ In the Belarusian dialectological tradition, the forms called here ‘adnumeratives’ have usually been referred to as vestiges of the Common Slavonic DUAL. See for example the quote by Hulickaja et al. (1992) in (§5.2.1.).

- (103) a. dv-a 'brata rob-ljatj
 two-NOM.M brother.GEN.SG work-PRES.3NONG
- b. dva 'bratɪ rob-ljatj
 two-NOM.M brother.ADNM work-PRES.3NONG
- c. dva bra'tɪ rob-ljatj
 two.NOM.M brother.NOM.PL work-PRES.3NONG
- 'Two brothers are working.'
- d. pjetj bra't-iv rob-ljatj
 five.NOM brother-GEN.PL/GRADNM work-PRES.3NONG
- 'Five brothers are working.'

Given the multiple configurations permitted, it is often impossible to tell whether numerals are governing the phrase (or showing head-like properties) or whether there is agreement. In any case, different solutions are needed to give account of the reality in West Polesian, and perhaps this also calls for a revision of some of the analyses for Russian.

5.5. Typological analysis of the adnumerative forms

In the previous sections, I had decided to analyse the adnumerative forms as a **NUMBER** value. I recognise that there are other possible ways of analysing it (not free from problems, either). But before jumping into other hypotheses, let us examine the consistency of this analysis from a typological perspective.

5.5.1. The ADNUMERATIVE as a **NUMBER** Value

I will start by picking up the hypothesis from the beginning:

H1: Adnumerative forms are a **NUMBER** value.

There is a considerable amount of linguists (working not exclusively with Slavonic languages) who are more inclined to analyse adnumerative forms as a **NUMBER** value (Bailyn & Nevins 2008, Igartua & Madariaga 2018, Nurmio & Willis 2016, Sims-Williams 1979). The resulting structure of features and their values is the following one:

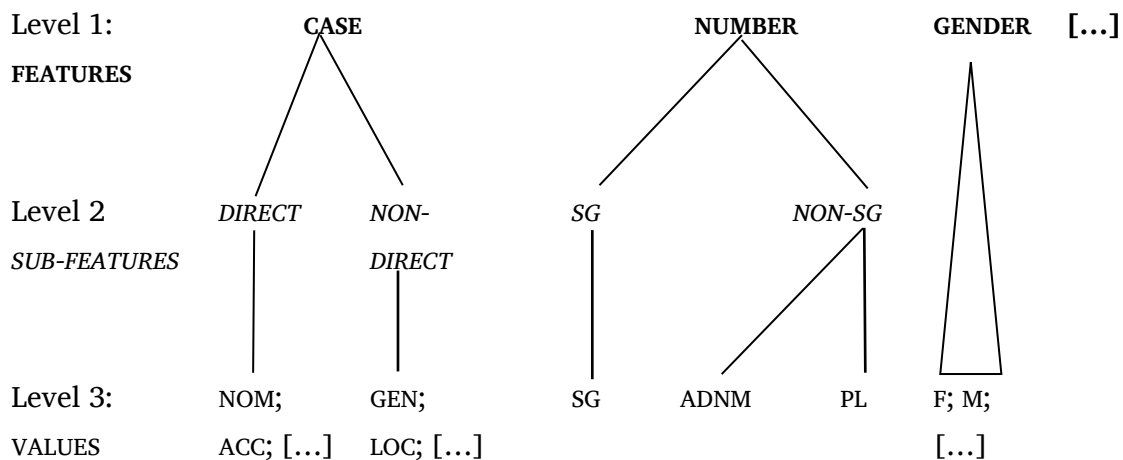


Figure 6 The ADNUMERATIVE as a NUMBER value

There is some good evidence that favours an analysis of West Polesian ADNUMERATIVE as a **NUMBER** value (H1):

First of all, we know that the ADNUMERATIVE derives etymologically from a DUAL **NUMBER** which was once more robust. This is also the case for Sogdian and Ossetian (Sims-Williams 1979, (forthcoming)), but as for Celtic languages (Nurmio & Willis 2016: 303), it has emerged in a different way, yet still very closely related to a **NUMBER** value distinction.

Secondly, the adnumerative forms have some semantically-defined basis, that is to say, they codify (or help codify) quantity, although not fully.

Thirdly, the adnumerative forms are in complementary distribution with the SINGULAR and PLURAL sub-paradigms of the DIRECT CASES.

Fourth, a much weaker argument is that some of the languages with some type of adnumerative, which I have described above, have very poor or no CASE morphology (synchronically), such as Bulgarian and Macedonian or the Celtic family. Notwithstanding, they all distinguish NUMBER (SINGULAR-PLURAL) consistently. So, it is easier to imagine an additional value in a feature that is already solidly established than to justify the need of a whole new feature for a value which appears very rarely. Yet, as I mentioned above, I recognise that all the languages described above did have a (more or less robust) CASE system in the past.

From the point of view of typology, particularly if we take a “Morphology-free syntax approach” (Baerman et al. 2005, Corbett 2012), positing the existence of additional (NUMBER) values is not only justified by the effects of the adnumerative forms on the syntax, but particularly by the existence of some phonologically dedicated ADNUMERATIVE forms. Now, even though I do not want to go deeper into syntax, there would be at least two sub-hypotheses deriving from H1 (and which both have been proposed for similar constructions in Russian, as I show in (§5.4.):

H1a: The ADNUMERATIVE is a NUMBER value available for other nominals (i.e. the underlying nominal morphology is richer than it appears at first glance).

H1b: The ADNUMERATIVE is a controller NUMBER only available for nouns, and the rest of the constituents in the NP agree with the material available to them.

H1a can be defended by assuming that those forms that are not phonologically dedicated (i.e. those that follow ‘the Russian system’, borrowing the forms of either the GEN SG or NOM PL for the NOM ADNM) are morphologically non-autonomous like ANIMACY.¹⁴⁷ Conversely, H1b, although descriptively perhaps less elegant, can be defended by saying that it is the best explanation for some phenomena which otherwise are very hard to account for (e.g. why the combination of a noun in ADNM + adjective in NOM PL is not permitted).

Both sub-hypotheses require a significant amount of assumptions and descriptive detail I do not intend to expand upon here, as I prefer to devote time to other hypotheses, but, before we proceed to the discussion about which feature value the adnumerative forms belong to, let us first explore how a proper (or canonical) NUMBER should look.

5.5.1.1. Canonical NUMBER

Corbett’s (2000) definition of NUMBER is quite strict. I will give more details on Canonical Typology (CT) in the next chapter (§6.5.)). This theoretical framework aims to find a base for comparing phenomena cross-linguistically. In particular, Corbett’s (2012) work focuses on feature values. Even though the framework of Canonical

¹⁴⁷ Although from a different theoretical framework, the analysis for Russian NumPs with lower numerals by Bailyn & Nevins (2008) is in line with this approach of enriching the morphology (i.e. extending the values of the ADNUMERATIVE to other constituents, at the expense of abundant syncretic cells) in order to explain the syntax of NumPs.

Typology, as such, is a later development; Corbett (2000) already presented several criteria for determining what a ‘well-behaved’ (or ‘canonical’ using the later terminology) (nominal) **NUMBER** value looks like. I have also added criteria for canonical features and their values from his later work (Corbett 2005, 2008, 2012, 2013).

Canonical criteria for nominal NUMBER	Violations
CRITERION 1 Adds semantic distinctions of NUMBER (individuation of items) to the system. ¹⁴⁸	· Alternative means of NUMBER marking (e.g. augmented stems). · Special discursive/semantic uses (honorifics, affectives, i.a.) (e.g. Icelandic).
CRITERION 2 Respects the Animacy Hierachy (Corbett 2000) (i.e. if available for any nominals, it will be for first person pronouns, then second person, etc.).	‘Second system’ (often in minor NUMBERS) (e.g. Maltese DUAL). ¹⁴⁹
CRITERION 3 Matches a typologically attested system (especially in terms of agreement).	Exceptional system (e.g. Baiso, Bezhta).
CRITERION 4 Semantically autonomous (either marks NUMBER /DETERMINATION; or denotes a quantity on its own).	GENERAL and UNDETERMINED NUMBER values (e.g. Basque).

¹⁴⁸ “We also looked in detail at systems which are unusual in terms of marking of number, yet which do not give new semantic distinctions. These were a useful reminder of the need to be clear whether claimed generalizations relate to number values or to the means of their expression, and more generally to be careful about comparing like with like” (Corbett 2000: 177).

¹⁴⁹ “Finally we looked at various languages, [...] which required us to extend the typology, primarily by introducing the notion of ‘second system’, which allowed for languages to have a type of number system operating for a lower part of the Animacy Hierarchy which would not normally be found as a top system” (Corbett 2000: 132).

Canonical criteria for nominal NUMBER	Violations
CRITERION 5 Morphologically autonomous (synthetic); expressed with a single word-form (Thornton 2019).	Periphrasis (e.g. expression of plurality by reduplication).
CRITERION 6 Available for all (quantifiable) nouns and pronouns.	· MINOR NUMBER (e.g. Hebrew DUAL). · Defectiveness (e.g. pluralia tantum nouns).
CRITERION 7 Obligatory marking.	Optional NUMBER marking (e.g. DUAL in Slovene and Ancient Greek).
CRITERION 8 Affects (syntactic or semantic) agreement.	Exclusively nominal NUMBER (NP agreement) (e.g. Lezgian).
CRITERION 9 Feature specifications on the controller and target match. ¹⁵⁰	· Default NUMBER. · Mismatch (small exception).
CRITERION 10 Its presence does not trump other NUMBER values.	Locally unmarked NUMBER: fossilised forms (e.g. plural of body parts in Polish).
CRITERION 11 It covers the same range as the rest of the NUMBER values in the language.	Multiple systems (top, second and bottom) (e.g. Yimas PAUCAL).
CRITERION 12 It covers the same range as the rest of the targets that mark it (e.g. the verb marks the same distinctions (SG-DU-PL) as the pronoun).	Conflated NUMBER (e.g. SG/DU conflation in Pame).

¹⁵⁰ “[T]he system of controller numbers and that of target numbers may be different [...] the controller may be outside the number system, giving rise to default number [...]. Or else the systems may be in harmony, but particular instances of controller and target may not coincide in number under certain circumstances” (Corbett 2000: 179).

5.5.1.2. West Polesian ADNUMERATIVE in the Light of Canonical NUMBER

Now that there is a base upon which to establish a definition of **NUMBER**, let us explore how or where the West Polesian ADNUMERATIVE fails to meet some of the important criteria.

CRITERIA 1 and 2: Adds semantic distinctions and respects the Animacy Hierarchy.

We can still believe that it adds semantic distinctions of **NUMBER** to the system, especially because of its etymology. However, there is no way to justify that it is not available for the top segments of the Animacy Hierarchy, which for Corbett (2000) is a crucial criterion. In this respect, we have to say that if West Polesian ADNUMERATIVE is a **NUMBER** value, it is part of a secondary system, where there is more flexibility for violations.

CRITERION 3: Matches an attested system.

Corbett himself also presents two counterexamples to this. Bezhta (Dagestanian) (Corbett 2012)[ISO 639-3: kap] has a minor PAUCAL **NUMBER**, but not a DUAL; even though, as a typological generalisation, a PAUCAL is not to be expected unless there is a DUAL.¹⁵¹ Baiso (Cushitic) (Corbett 2000) [ISO 639-3: bsw] can have a PAUCAL without a DUAL; yet, according to Corbett (2000, 2012), Baiso's **NUMBER** system is far more robust and complex than the one found for West Polesian.

¹⁵¹ Bezhta's PAUCAL is not similar to the West Polesian ADNUMERATIVE, because Bezhta's PAUCAL is only restricted to a very limited list of nouns, which is *per se* semantically restricted.

I have no evidence of any description of a language in which the ADNUMERATIVE exists as a full/properly distinguished and syntactically independent NUMBER value. However, it must also be said that there is not much work available on the adnumerative, so this should not be a definitive reason to discard the hypothesis. For this reason, I have tried to relate the adnumerative to some well attested or better known NUMBER values, to see which of them suits better:

· A GENERAL NUMBER (a.k.a. TRANSDUMERAL):¹⁵² According to Corbett (2000) a GENERAL number denotes “an unspecified amount of X” when it stands on its own. The particularity of this NUMBER value is that it attaches to any numeral form, whilst the form is beyond the SINGULAR-PLURAL distinction. One of the advantages of describing the ADNUMERATIVE as a general NUMBER value is that it displays an ‘unmarked’ agreement in the past (assuming that the noun is the head, instead of the numeral), and that there are different agreement possibilities when followed by adjectives. The general numeral is the closest attested system to the adnumerative, but there are two main differences:

- a) The ADNUMERATIVE imposes many semantic restrictions (only nouns from ‘two’ to ‘four’ or derived), whereas canonical GENERAL NUMBERS /TRANSDUMERALS should be available for any cardinal numeral, except, perhaps ‘one’.
- b) The ADNUMERATIVE cannot stand on its own, it is syntactically non-autonomous.

¹⁵² “Acquaviva (2006: 1868-1869, 2008: 188) terms the Irish counting forms “transnumeral” and considers them inherently plural forms, arguing that they are “semantically unit counters” which function as classifiers” (Nurmio & Willis 2016: 297). Nevertheless, the mixture of PLURAL and TRANSDUMERAL (which according to this it would be a subclass of PLURAL) suggests that most likely Acquaviva’s understanding of TRANSDUMERAL is very different from what is being presented here.

· A PAUCAL. Igartua & Madariaga (2018) and Madariaga & Igartua (2017) postulate that the Russian analogous form (used in NumPs with ‘two’, ‘three’, ‘four’ and ‘both’; i.e. allegedly GEN SG) is a "paucal construction", which is a common term in the Russian Minimalist tradition (Nerea Madariaga, p.c.).¹⁵³ In Madariaga & Igartua’s (2017) article they make clear that their understanding of PAUCAL differs from Corbett’s (2000) concept (theirs being more inclusive than Corbett’s). If we stick to the narrow definition of PAUCAL, there are two major objections to this: one is a syntactic, the other one is semantic. Firstly, the ADNUMERATIVE is syntactically non-autonomous, which, to be fair, we will see recurring in all the comparisons with other NUMBER values as well. And secondly, from the point of view of semantic typology, most PAUCAL denote an approximate quantity that is not very big, but they rarely mean something like ‘exactly from two to seven’. Conversely, West Polesian ADNUMERATIVE is *exclusively* restricted to numerals ‘two’ to ‘four’ (and their derived forms). That is to say, the amount of entities that the ADNUMERATIVE denotes is very restricted or very well-defined. Finally, although admittedly this is a far weaker point, PAUCAL number is typologically rare, and geographically concentrated in Austronesian languages, rather than Indo-European.

· A DUAL. We know that the ADNUMERATIVE derives from an eroded DUAL, and that it is marked in nouns headed by the numeral ‘two’ (and not so clearly, ‘both’). However, besides its lack of syntactic autonomy, the most obvious reason why it cannot be a DUAL is that it also appears after other numerals. It is unwise to propose a TRIAL and a

¹⁵³ I understand that by adding the noun "construction", Madariaga & Igartua (2017) suggest that this PAUCAL is syntactically non-autonomous.

QUADRAL which are all identical in every respect (especially taking into account that the QUADRAL is extremely rare typologically, and the fact that it is not available for FIRST or SECOND PERSON pronouns discredits any further analysis).

I summarise all the possible analyses of the ADNUMERATIVE as a NUMBER value, balancing evidence for (thus, it is a canonical NUMBER value) and against it (fails to meet the requirements of the canon for a NUMBER value) in Table 24 (below).

Thus, the ADNUMERATIVE does not fit into any of the well-attested more robust NUMBER systems, and in this respect it is non-canonical.

CRITERION 4: Semantic autonomy.

The ADNUMERATIVE does not have full meaning on its own. That is to say, forms like [ADNM] *'bratı* ‘brother(s)’ do not denote on their own ‘a few (more than one, less than five) brothers’. This seems to be shared in common by all adnumerative forms cross-linguistically. For example, Nurmio & Willis (2016) make the following remark on the Welsh ADNUMERATIVE:

“While it is clearly related to the category of number, it is not itself a number category under Corbett’s (1996, 2000) approach, since it has no independent semantic value: a numerative form on its own does not mean ‘some defined number of x’. For this reason we will term it a minor category, rather than a minor number¹. [...] This does not present any significant conceptual problem, since minor categories exist within other areas, such as gender and case, for instance, the Russian second locative case (Brown 2007)” (Nurmio & Willis 2016: 302).

Table 24 Hypothetical equivalents of the ADNUMERATIVE

NUMBER value	For (thus, canonical)	Against (thus, non-canonical)
GENERAL NUMBER/ TRANS- NUMERAL	-Its main function is counting or it does not specify a particular amount of entities. -Displays an UNMARKED agreement in the past, and multiple agreement possibilities when followed by adjectives.	-It is very restricted semantically (i.e. it cannot appear with any numeral). -Syntactically non-autonomous.
PAUCAL	-It only appears with lower numerals, and it is more inclusive than the DUAL.	-No independent meaning on its own. -Syntactically non-autonomous. -Areally/Typologically unlikely.
DUAL	-Etymologically derived from a DUAL. -It is used to indicate ‘two’ (more than one and NON-PLURAL) entities.	-It also involves numerals ‘three’ and ‘four’. -Syntactically non-autonomous.

CRITERION 5: Synthetic, expressed with a single word form.

The ADNUMERATIVE is realised as a suffix, combined with stress, but does not need any additional suffixes or words to mark it (it is morphologically autonomous). Thus, according to this criterion, the ADNUMERATIVE meets the canon.

CRITERION 6: Available for other nominals.

If we disregard H1a, this criterion is particularly tricky in this respect. On the one hand, there are no ADNUMERATIVE forms available for the pronouns. We could justify that gap by a secondary system, which is incomplete due to the scarcity of items affected; i.e., a MINOR NUMBER category. However, the ADNUMERATIVE is morphologically robust enough to qualify for a MINOR NUMBER category, since potentially all the countable nouns have or can have one. Having said this, Corbett (2000) points out that Avar has a MINOR PAUCAL category; however, the term *minor* does not seem very appropriate to describe it, as more than eighty nouns have it.¹⁵⁴ In this respect West Polesian ADNUMERATIVE is non-canonical, but we see that it is not alone.

CRITERION 7: Obligatory marking.

Even though there are alternative forms to the dedicated ADNUMERATIVE form, the use of an adnumerative is always mandatory when headed by a lower numeral. Thus, in this respect, the West Polesian ADNUMERATIVE does meet the canon.

CRITERIA 8 & 9: Affects agreement and there are no mismatches of values.

Regardless of the hypothesis or sub-hypotheses we eventually take, there is no doubt that it does have an effect on agreement (and thus it is canonical in this respect). Whether or not there are mismatches of feature values depends a lot on the sub-

¹⁵⁴ So, according to Greville G. Corbett (p.c.) ‘minor’ should be understood almost as a secondary system and not the size of the class. That is to say, that specific NUMBER value would be defective in the ANIMACY HIERARCHY, like other well-behaved NUMBER values in the language.

hypothesis we want to follow. Nevertheless, most data point to the use of a default **NUMBER** value (usually **NEUTER SINGULAR** or **PLURAL**) in domains where the **ADNUMERATIVE** appears, but it is also very likely that the default verbal marking comes from the NumP instead. Hence, depending on our final analysis, the adnumerative can be a canonical feature value or not in this respect.

CRITERION 10: Does not trump other **NUMBER** values

The fact that in **OBLIQUE CASES** the **ADNUMERATIVE** shares its paradigm with the **PLURAL** sub-paradigm suggests that perhaps the presence of the **ADNUMERATIVE** trumps **PLURAL**. By contrast, the fact that NumPs containing an **ADNUMERATIVE** can trigger **DEFAULT** verbal agreement in the **PAST** counterbalances this. So, since there are no clear indices of the **ADNUMERATIVE** overtaking a particular **NUMBER** value, in this respect it is canonical.

CRITERIA 11 & 12: Covers the same range as the rest of the **NUMBER** values in the language and the targets that mark it.

Excluding uncountable nouns, there is no shadow of a doubt that **SINGULAR** and **PLURAL** values are available for virtually all the Animacy Hierarchy. Conversely, there are no pronouns that mark **ADNUMERATIVE**, thus, the **ADNUMERATIVE** can only be a second system, and thus non-canonical.

5.5.2. The **ADNUMERATIVE as a **CASE** value**

Some well-known linguists (Corbett 2008, Mel'čuk 1985, Zaliznjak 1973, 2002), notably working on Russian, have analysed the Russian **ADNUMERATIVE** as a **CASE** form.

Following the descriptive/analytic methodology from the beginning, let us reformulate the analysis as a hypothesis:

H2: The ADNUMERATIVE (or adnumerative forms) is a CASE value.

The structure of the ADNUMERATIVE derived from H2 is to be the following (Figure 7):

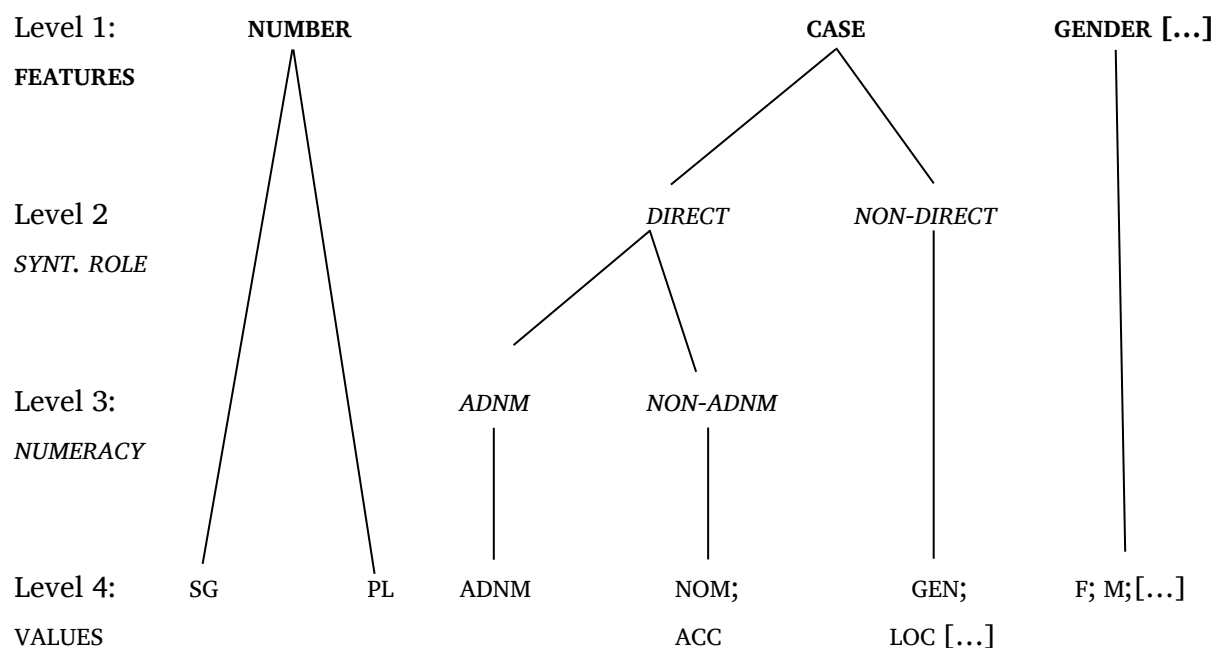


Figure 7 The structure of features and their values predicted by H2.

This hypothesis predicts noun inflectional paradigms like the one in Table 14.a. It is still unclear whether the ADNUMERATIVE belongs to the SINGULAR sub-paradigm (e.g. like the VOCATIVE, only available for the SINGULAR), the PLURAL sub-paradigm; or whether it is independent from NUMBER constraints. The way the ADNUMERATIVE is presented in Table 14.a favours this last approach. Yet, this outlying is due to mere

practical reasons (trying to keep the most neutral of all the representations), but it does not presuppose that it is the ultimate or the best analysis.

There are some good arguments to believe so:

Firstly, the ADNUMERATIVE behaves like a CASE value in as much as it codifies the syntactic relation between a numeral and a noun, though the function of the ADNUMERATIVE seems ostensibly redundant as it only indicates that the heading numeral is a lower one.

Secondly, the lack of ADNUMERATIVE CASE for other targets could explain the competition of forms as in (73).

Thirdly, this hypothesis (H2) could explain why phrases containing numerals can sometimes override ANIMACY. If a noun followed by a LOWER NUMERAL must stand in ADNUMERATIVE CASE, then ANIMACY cannot force it to switch to a different CASE (ACC = GEN).

Fourthly, from a morphological or morphophonological perspective, when there is disunity regarding the stem or the type of stress between the SINGULAR and PLURAL sub-paradigms, the ADNUMERATIVE aligns with the SINGULAR in all the instances analysed so far. I have already indicated in (§5.2.2.) that parasitically, the ADNUMERATIVE can be derived from the OBLIQUE SINGULAR stems:¹⁵⁵

¹⁵⁵ Following the Classical tradition, it is tempting to say that it originates from the GENITIVE, although all the paradigms analysed so far show that it can be derived from any OBLIQUE SINGULAR form.

(104)

(a) 'horse'	SINGULAR	PLURAL
NOM	kinj	'konɪ
ACC	ko'nja	'konɪ
GEN	ko'nja	'konej
DAT	ko'nju	ko'nj-am -om ¹⁵⁶
INS	ko'njom	'kinjmi
LOC	(na) ko'nji	(na) 'konjax
ADNM	ko'nji	

(b) 'child'	SINGULAR	PLURAL
NOM	dɪ'tja	'djitɪ
ACC	dɪ'tja	de'tej
GEN	dɪ'tjati	de'tej 'djɪtok
DAT	dɪ'tjati	'djɪtj-am -om
INS	[n.d.]	[n.d.]
LOC	[n.d.]	[n.d.]
ADNM	dɪ'tjati	

In (104) a., b. we see that the ADNUMERATIVE form follows the stress pattern of the SINGULAR sub-paradigm, but then the DATIVE PL in (104) a. and the ACCUSATIVE/GENITIVE PL in (104) a., b. for no apparent reason have the stress on the suffix (like the SINGULAR). In regard to stem alternation, the whole SINGULAR sub-paradigm has a palatalised ending in (104) a. and non-palatalised in (104) b., which is replicated by the ADNUMERATIVE form; whereas, when it comes to the PLURAL sub-paradigm, there is inconsistency again. Thus, using morphological patterns as an argument for a certain analysis (ADNUMERATIVE as a CASE of the SINGULAR sub-paradigm) does not prove to be a solid criterion.

Certainly, there are also some **major concerns** against such an analysis (H2):

Firstly, the ADNUMERATIVE would be in complementary distribution with DIRECT CASES,¹⁵⁷ as we have just seen. However, this is also problematic for such an analysis. In spite of their complex internal syntactic structure, the NPs where the ADNUMERATIVE appears keep behaving like regular NPs in terms of their primary syntactic function.

¹⁵⁶ No data available in my notes. This form has been extracted from Paulus (1970).

¹⁵⁷ As I have previously shown NumPs with adnumeratives are not necessarily affected by the animacy constraints (ACC PL = GEN PL) in at least some varieties, or just simply adnumeratives override animacy, and so the difference between animate and inanimate is neutralised.

There is no doubt that they fulfil the role of arguments of the verb (subject/object) as other NPs marked by DIRECT CASES (especially because verbs in West Polesian always require the subject to be in the NOMINATIVE). Thus, there are two possible solutions for this:

- a) Postulating that the ADNUMERATIVE is a ‘COMBINED CASE’, in terms of being able to mark phenomena happening at different levels: SYNTACTIC ROLES (at the level of feature values) and NUMERACY (in the structure proposed in (Figure 7), underneath SYNTACTIC ROLES). From the point of view of typology it seems hard to imagine this, especially because different layers are being mixed.
- b. Changing the structure proposed above in order to include the ADNUMERATIVE under each DIRECT CASE (consequently generating two CASE values (*de facto*) instead of one) (Figure 8):

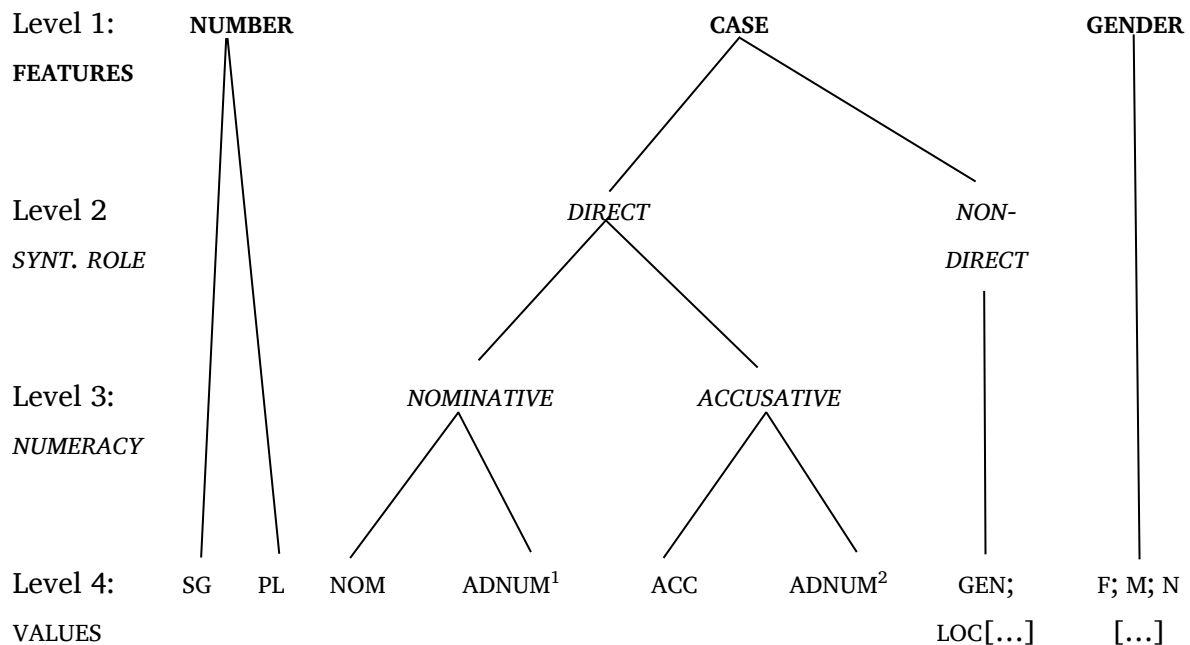


Figure 8 Structure of features and their values predicted by (H1) with two ADNUMERATIVES as CASE values

Nevertheless, proposing two ADNUMERATIVE SUB-CASES, instead of one, seems a far more problematic solution than the previous one. The main reason is that we have to propose that both CASES are morphologically (and thus, phonologically) absolutely syncretic, which happens very rarely in typology (Corbett 2012: 182) (and in this instance other analyses are possible).

Secondly, some may argue that there seems to be an ADNUMERATIVE PLURAL (which I call GREATER ADNUMERATIVE) counterpart (e.g. B6: *traxto¹rv/ej* tractors) for higher numerals, which Mel'čuk (1985) also documents appearing with a handful of nouns in Russian. I have written extensively on the differences between the two forms in (§5.3.1.3.) and (§5.3.2.2.). Therefore, even though there are cues to establish a correlation between the two adnumerative forms, this correlation is not entirely symmetrical, and thus not an entirely valid argument to be used in favour of H2.

Finally, related to the former point, the (LOWER) ADNUMERATIVE would be a CASE value which does not cross-cut with NUMBER (at least, not fully).¹⁵⁸ Corbett (2012) remarks that the Russian ADNUMERATIVE is only available for the SINGULAR sub-paradigm, which he also finds unsettling, and which, at least at first glance, can also apply to the West Polesian ADNUMERATIVE:

“If it is indeed a case value, it is unusual in being available only in one number value (for any given item). But then, Russian numerals typically govern a case value and require a particular NUMBER value on the noun,

¹⁵⁸ However, according to Matthew Baerman (p.c.) this is something to be expected, because the form is governed by quantifiers.

and so the ADNUMERATIVE fits into this pattern. Are we dealing with government?” (Corbett 2012: 261).

Nonetheless, West Polesian ADNUMERATIVE does not clearly belong to the SINGULAR or PLURAL sub-paradigm (as it can be thought of the Russian ADNUMERATIVE).¹⁵⁹ In addition to this, Nurmio & Willis (2016) show reluctance to analyse the ADNUMERATIVE as a CASE value. They understand that Corbett’s (and above-mentioned authors advocating for this in Russian) “[...] motivation for treating forms after numerals as a CASE form comes from syncretism with genitive forms used elsewhere” (Nurmio & Willis 2016: 304) in Russian.

Hence, we can see that analysing the adnumerative as a CASE value is no less problematic. Corbett (2008), who advocates such an analysis, admits this is also troublesome: “The adnumerative seems to be on the extreme edge of what could be included as a case value” Corbett (2008: 19). But in order to take our critique to the end, let us establish what well-behaved CASE values should look like, as we have done for NUMBER.

5.5.2.1. Canonical CASE

It has been relatively easy to define a base upon which to build the canonicity criteria for NUMBER, whilst doing the same thing for CASE is a lot harder, or even impossible. Firstly, for NUMBER there are some well-defined minimal criteria (particularly

¹⁵⁹ In fact, other nouns in Russian such as *ča'sa* [ADNM] ‘hour(s); clocks’, follow the stress pattern of the plural sub-paradigm (instead of the one for the singular ([GEN SG] *'časa*; [NOM PL] *ča'sy*) except for the [SECOND LOCATIVE] *ν ča'su*). This should also call us to question whether there is any evidence that Russian ADNUMERATIVE belongs to the SINGULAR sub-paradigm.

respecting the Animacy Hierarchy; and morphosyntactic autonomy) which are followed quite strictly by most languages. But CASE values are not subject to any model or hierarchy as NUMBER values are.

Secondly, NUMBER had a well-defined semantic basis which helps to individuate or quantify entities. By contrast, CASE is often used as a jumble where heterogeneous marks with very different functions are put: from purely syntactic (e.g. ERGATIVITY), to quantificational (e.g. PARTITIVES) or temporal-spatial (e.g. ADLATIVE).

And thirdly, one of the requirements of NUMBER values was that they must match an existing/attested system (of NUMBER values), whereas with CASE there is not such a restriction. New CASE names or forms are constantly being proposed in language descriptions, and their possibilities seem unlimited. There is a relative freedom for the linguist to be 'creative' with the tags for CASE values and their syntactic functions (i.e. it is the most 'open class' of morphosyntactic features). Of course, this also poses a major impediment when comparing specific CASE values cross-linguistically, as the functions and domains of use can vary drastically from one language to another.

Having presented the major handicaps to elaborate this typology, let us proceed to set a list of criteria for canonical CASE values, admitting that these are less strict than for NUMBER values. The remaining canonicity criteria, which are general to all feature values, will appear in (§5.5.3.).

CRITERION 1: Canonical CASE values mark a relation between at least two constituents.¹⁶⁰

When talking about CASE marking, these constituents are usually the head and the target. When it comes to the ADNUMERATIVE, its main function is to mark a relation between the numeral and the noun. Thus, in this respect, the ADNUMERATIVE meets the canon.

CRITERION 2: Canonical CASE values fulfil a (semantically justified) syntactic role.

We do not expect to find a CASE marker on the noun ‘sea’ that simply encodes ‘being blue’ in a sentence like: *We went to the sea/ The sea swallowed our boat*. When it comes to the ADNUMERATIVE, this is not an easy question. On the one hand, we can say that the ADNUMERATIVE is fulfilling a quantificational role (but, in this respect, it is also closer to the functions of NUMBER values). However, on the other hand, the ADNUMERATIVE is only indicating that the preceding numeral is a lower one (and perhaps, that it is in a DIRECT CASE). Hence, the ADNUMERATIVE fails to meet this canonicity criterion.

CRITERION 3: Canonical CASE values only mark one syntactic role at a time.

According to Oliver Bond (p.c.) in order for a CASE value to be canonical, it needs to be predictable. So, if a CASE is marking two values at the same time it is impossible to predict the other value. For example, a single canonical CASE value cannot mark LOCATIVE and GENITIVE at the same time. When it comes to the ADNUMERATIVE, this is a

¹⁶⁰ “[...] case morphology, though governed by rules that refer to the syntax, constitutes an independent level of linguistic analysis, whose function is to mediate between the categories and configurations of syntax and the actual forms supplied by the morphology” (Pesetsky 2013: 5).

major objection, as it would be marking NOMINATIVE or ACCUSATIVE as well as ADNUMERATIVE at the same time with a single suffix. For this reason, the ADNUMERATIVE does not meet the canon in this respect.

CRITERION 4: Canonical CASE values are orthogonal to other well defined feature values (most notably GENDER and NUMBER).

The ADNUMERATIVE is available for (countable and inflectable) nouns from any inflectional class. However, the ADNUMERATIVE does not behave like any other CASE value in the paradigm, in that it is unclear which NUMBER value it is available for. As a result, the ADNUMERATIVE does not meet this canonicity criterion.

CRITERION 5: Canonical CASE values are not affected by other feature values.

Corbett (2012) describes how Latvian prepositions can affect CASE assignment depending on the NUMBER. ANIMACY is not a feature value in itself (although Corbett (1991) treats it as a SUB-GENDER in the Slavonic family, so it would be a SUB-VALUE), but it can have an effect on the ADNUMERATIVE (although it can also override it). So, in this respect, the ADNUMERATIVE is a less canonical CASE value.

CRITERION 6: Canonical CASE values are realised within a single word form (e.g. they are synthetic).

As had been already proposed for NUMBER values (*supra*), the marking must be synthetic in order to be canonical (based on Thornton (2019)). Because the ADNUMERATIVE is realised like any other CASE/NUMBER value, the ADNUMERATIVE meets this canonicity criterion.

5.5.3. The West Polesian ADNUMERATIVE from the point of view of CT

We have seen that from a typological point of view it is hard to determine whether the West Polesian ADNUMERATIVE is a CASE or a NUMBER value. Corbett (2013: 48) recognises that even well-known features such as CASE, NUMBER or PERSON often show different degrees of canonicity:

“[...] yet we should recognise that feature systems are more complex than that. Features vary: (a) according to how well founded they are, and (b) in how they are distributed across the lexicon. To analyse this difficult area, I start from an idealized view, and then plot the deviations from that ideal. In other words, I take a canonical approach” (Corbett 2013: 48).

Once to this point, I will analyse the West Polesian ADNUMERATIVE following the criteria for canonical features and their values set in Corbett (2011, 2012, 2013).

Principle I: *Features and their values are clearly distinguished by formal means (and the clearer the formal means by which a feature or value is distinguished, the more canonical that feature or value).*

Irrespective of the analysis we eventually take, West Polesian NUMBER and CASE are very well-defined features in other parts of the system. Their values tend to be well defined for the main five-six CASES as well as the SINGULAR/PLURAL; whereas the ADNUMERATIVE form(s) do not behave anything like that, as we are going to see.

CRITERION 1: Canonical features (and their values) have dedicated forms (autonomous).

I have already mentioned that the suffix for the ADNUMERATIVE is *-I* regardless of the inflectional class. But, as with many other forms in the paradigm, having a dynamic stress often helps to disambiguate forms that are otherwise syncretic (from a pure segmental perspective). Yet, according to the data gathered so far, the ADNUMERATIVE of inflectional class I nouns is always syncretic with the GEN SG as both suffixes are phonologically identical and the ADNUMERATIVE always follows the stress pattern of the OBLIQUE SINGULAR CASES. Of course, some people may object that there is no dedicated ADNUMERATIVE form for inflectional class I; but, then we would have to propose very complex syntactic rules to explain why the ADNUMERATIVE only appears with inflectional class II and III nouns, which is certainly problematic for a “Morphology-free-syntax” approach (Baerman et al. 2005) and it bears the same suffix as the rest. Moreover, we would have to explain why the rest of the constituents in the sentence do not take the forms we expect from straightforward NOMINATIVE or PLURAL subjects (i.e. NEUTER SINGULAR verbal agreement, GENITIVE PLURAL NP agreement). Thus, if we accept that the link between the inflectional class I ADNUMERATIVE and the GENITIVE SINGULAR (as well as many NOMINATIVE PLURAL forms) is, at least synchronically, purely accidental, West Polesian ADNUMERATIVE meets the first criterion, but *only* for DIRECT CASES.

CRITERION 2: Canonical features and their values are uniquely distinguished across other logically compatible features and their values.

Firstly, regardless of the hypothesis or analysis we adopt (CASE or NUMBER), the adnumerative fails to be canonical in both respects, as already from the starting point,

it is not clearly distinguished from the features of **CASE** and **NUMBER**. Furthermore, the **ADNUMERATIVE** neutralises the differences between inflectional classes, as all the nouns have the same suffix for the **ADNUMERATIVE -i**. Yet this is not extraordinary, since many **CASE** suffixes in the **PLURAL** (**OBLIQUE CASES**, excluding the **GENITIVE**) are identical for all nouns (that have a **PLURAL** sub-paradigm) in many Slavonic languages. Moreover, in some idiolects **ADNUMERATIVES** (and numerals, in general) trump **ANIMACY**.

Secondly, if we analysed the **ADNUMERATIVE** as a **CASE**, the **ADNUMERATIVE** would be in conflict with **DIRECT CASES**, since feature combinations of the type **NOM SG/PL + ADNUMERATIVE** in one single noun are impossible. The noun has to be either in **NOMINATIVE/ACCUSATIVE** *or* **ADNUMERATIVE**, even though the **NumP** with the **ADNUMERATIVE** is fulfilling the syntactic role of a **NOMINATIVE/ACCUSATIVE**. In addition, the **ADNUMERATIVE** cannot be easily attached to the **SINGULAR** or **PLURAL** sub-paradigm, as happens with other marginal/defective **CASE** values.

Finally, if we analysed the **ADNUMERATIVE** as a **NUMBER** value, it would also be non-canonical as it is restricted to **DIRECT CASES**, which is suspicious (unless we adopt the approach of enriching the morphology (H1a)). Notwithstanding the approach or hypothesis we eventually take, the **ADNUMERATIVE** fails to meet this canonicity criterion as a feature value.

CRITERION 3: Canonical features and their values are distinguished consistently across parts of the speech.

On the one hand, if we assumed the **ADNUMERATIVE** exists only for nouns (as **CONTROLLER NUMBERS** or defective **CASE** forms), the **ADNUMERATIVE** would be completely non-canonical in this respect. The apparently illogical agreement patterns appearing

in the presence of NumPs containing an ADNUMERATIVE point out a gap in the rest of the system to accommodate the ADNUMERATIVE. According to Corbett (2013: 50) this is not surprising for CASE, which because of its idiosyncrasy is never available for controllers/governors. However, it becomes more problematic if we want to analyse the ADNUMERATIVE as a NUMBER value.

Depending on the analysis we take (CASE or NUMBER) the ADNUMERATIVE would be either “a non-autonomous CASE/NUMBER value”, or “a dependent controller CASE/NUMBER value”, following the terminology in (Corbett 2012: 171),¹⁶¹ which realises its values “only through syncretic forms” (i.e. borrowing from other parts of the paradigm).

On the other hand, if we believed in the above-mentioned sub-hypothesis (H1a), which argued for a richer morphology of the adnumerative forms, the ADNUMERATIVE would still be non-canonical in this respect. The reason is that although there would be adnumerative forms available for the agreeing targets, these would not distinguish the ADNUMERATIVE from the GREATER ADNUMERATIVE/GENITIVE PLURAL as nouns do.

CRITERION 4: Canonical features and their values are distinguished across lexemes within relevant parts of the speech.

West Polesian ADNUMERATIVE is canonical in this respect. The vast majority of nouns can have an ADNUMERATIVE form (with the exceptions of *singularia tantum*, and the like), which is often homophonous with other cells. Against any first predictions, this canonicity can actually present more challenges for the analysis because if the ADNUMERATIVE was a

¹⁶¹ However, Greville G. Corbett (p.c.) dislikes the use of this last form (especially “dependent target”) for this phenomenon.

‘minor value’ (Corbett 2012: 185), as it could be in Russian (with just seven nouns), it would be easier to apply the criteria or predictions for MINOR NUMBERS in (Corbett 2000) (e.g. Hebrew and Maltese DUALS). That is to say, having a smaller sample of instances results in an analysis which is more permissive with irregularities.

Principle II: *The use of canonical morphosyntactic features and their values is determined by simple syntactic rules.*

CRITERION 5: The use of canonical morphosyntactic features and their values is obligatory.

For most (if not all) speakers the dedicated ADNUMERATIVE form can be replaced by the GENITIVE SINGULAR (looking-like) form, and for many, even by the straightforward NOMINATIVE PLURAL. Consequently, we must admit that there is a competition of forms or constructions, which make the ADNUMERATIVE less canonical as a feature value in this respect (like the VOCATIVE in West Polesian).¹⁶²

CRITERION 6: Canonical use of morphosyntactic features and their values does not admit syntactic conditions.

The ADNUMERATIVE fails to meet the canon in this respect. As I have mentioned before, there is lots of variation across speakers and even within the speakers in terms of the forms that can replace or compete with the ADNUMERATIVE (GEN SG and NOM PL). There is at least one condition that always alters syntactic agreement (a), and another condition that seems more a tendency than a rule (b):

¹⁶² However, the marking of CASE and NUMBER is always obligatory, and thus CASE and NUMBER are still canonical in this respect.

- a) If the numeral appears after the noun, the noun will take the form that corresponds to the higher numerals (i.e. GRADNUM), and the meaning of the numeral will be no longer exact, but approximate (105) (§4.2.3.).¹⁶³

(105) a. (B18.4E 00:37)

dva 'kɪɪ
two kilo.ADNM
'Two kilos'.

b. (B18.4 00:33)

nu, ʃtʃo ʃ, Xonj-e, **kɪl** **dva** bud-e
so what PART Xonja-VOC kilo.GEN.PL/GRADNM two.NOM.N be.FUT-3SG
'So what, Xonja, there'll be around two kilos.'

It is interesting to remark that when the NumP indicates an approximate quantity (as in (105) a.) the target verb agrees in SINGULAR (and not in PLURAL, as we would expect).

- b. The noun has to immediately follow the numeral, otherwise the syntactic agreement can change (as in (105) b.).

In any case, the ADNUMERATIVE fails to meet the canon, as it requires a certain type of numeral phrase to justify its presence. Otherwise, its use is ungrammatical.

CRITERION 7: Canonical use of morphosyntactic features and their values does not admit semantic conditions.

The ADNUMERATIVE fails to meet the canon in this criterion as well. This criterion is difficult to analyse. On the one hand, since the type of numeral (i.e. only lower numerals) that can be governing the numeral phrase (or agreeing with the noun) is

¹⁶³ So far I have only documented one exception, and it was a suppletive form.

very restricted, the condition is purely semantic. But on the other hand, Corbett's (2012: 193) understanding of condition goes beyond that. That is to say, certain numerals govern ADNUMERATIVE, as certain prepositions govern a specific CASE (Greville G. Corbett, p.c.), but the use of the ADNUMERATIVE will not be altered if the sentence is negated (i.e. polarity) or the noun is INANIMATE, for example. So, in this respect, the ADNUMERATIVE meets the canonicity criterion.

CRITERION 8: Canonical use of morphosyntactic features and their values does not admit lexical conditions from target or governee.

The ADNUMERATIVE meets this criterion because what triggers the corresponding form is the governing numeral phrase.

Now, it is important to mention that for a handful of words, particularly in the variety of Bahdanaŭka, there is a very intriguing phenomenon. The nouns for 'brother' and 'son' seem to have a preference for changing their form depending on the numeral governing them; i.e. [2] *'bratɹ*; [3] *bra'tɹ*; [4] *'bratɹ* 'brothers'. After multiple tests, I cannot say that the heading numerals govern a specific form, because there is a decent amount of versatility with the phonological forms of the ADNUMERATIVE. Yet, certain speakers have reported a preference for such a distribution, which, again, is not entirely consistent, but common, indeed. Interestingly enough, I have only properly attested this with the nouns 'son' and 'brother'; i.e. no other VIRILE nouns have such a distribution.¹⁶⁴

¹⁶⁴ I observed a similar instance with certain infl. class III nouns in the speech of Tor1, who insisted on that distribution. Nevertheless, I ran a similar questionnaire including such nouns a year later and she only provided the regular or 'expected' distribution.

CRITERION 9: Canonical use of morphosyntactic features and their values does not admit lexical conditions from the controller / governor.

According to Corbett (2012: 218), Russian ADNUMERATIVE, and thus, West Polesian ADNUMERATIVE “is non-canonical here, since the few items that govern it have an additional condition: the governor must itself stand in the NOMINATIVE (or the ACCUSATIVE identical to the NOMINATIVE, as occurs when the governed noun is inanimate).”

CRITERION 10: The use of canonical morphosyntactic features and their values is sufficient (they are independent).

None of the adnumerative forms analysed are syntactically autonomous. As I stated in (§5.2.) adnumerative forms are a strictly **syntactically bound form** in essence. They always require a heading numeral phrase (or quantifier) in order to appear as they lack meaning on their own. As we have previously seen, this has been the major criterion that disqualifies the ADNUMERATIVE from possible analyses of different NUMBER values.

Having said this, let us turn our attention to the LOCATIVE,¹⁶⁵ which is a far more robust feature value (or better established) than the adnumerative forms. The LOCATIVE is also well known for being syntactically non-autonomous in most Slavonic languages. The fact that it can only appear governed by a preposition has led certain grammar traditions to refer to it as PREPOSITIVE/PREPOSITIONAL. Corbett (2008, 2012) and Zaliznjak (1973, 2002) point out other less canonical minor CASE values, among which the

¹⁶⁵ By LOCATIVE (a.k.a. PREPOSITIVE/PREPOSITIONAL particularly in the Russian descriptive tradition) I refer to the primary or most robust locative form; i.e. the one that is available for virtually every inflecting noun, and not what is known as the SECOND LOCATIVE (Brown 2007).

INCLUSIVE (Corbett 2008: 23-26) also requires a heading preposition. Thus, the ADNUMERATIVE is not the only feature value failing to meet the criterion in West Polesian.

Principle III: *Canonical morphosyntactic features and their values are expressed by canonical inflectional morphology.*

For the remaining instances in the language, we know that both NUMBER and CASE are quite well defined. The number of CASE values can be tricky in West Polesian, as the ACCUSATIVE is a non-autonomous form (except for the FEMININE SINGULAR), as well as the DATIVE/LOCATIVE (except for the PLURAL) in many idiolects (see (§3.2.1.)).

When it comes to NUMBER, if we disregard the particular agreement NumPs trigger on the noun, and lexically specified nouns (including *singularia tantum*, *pluralia tantum*, uncountable or mass nouns and indeclinable nouns, if there are any),¹⁶⁶ it is realised quite canonically.

I have summarised the results following a binary approach in Table 25.

¹⁶⁶ So far, I have not been able to identify any indeclinable nouns in the corpus. The only potential noun (because of its indeclinable cognate in Russian) I have found is '*radio*', which appeared inflected in a text (B20)“ [...] *radio* ni buło” ‘there were no radios’. I have also overheard speakers from other regions (Sičyv) using the same form. Yet, unfortunately, I forgot to elicit the ADNUMERATIVE of such nouns.

Table 25 Summary

Criteria	ADNUMERATIVE
CRITERION 1: Autonomy	+
CRITERION 2: Distinguished across features/ values	–
CRITERION 3: Distinguished consistently across parts of speech	–
CRITERION 4: Distinguished consistently across lexemes	+
CRITERION 5: Obligatory	–
CRITERION 6: No extra syntactic conditions	–
CRITERION 7: No extra semantic conditions	+
CRITERION 8: No lexical conditions from target or governee	+
CRITERION 9: No lexical conditions from controller / governor	–
CRITERION 10: Independence	–

The results obtained after this general analysis do not seem particularly optimistic.

The ADNUMERATIVE only meets 4 out of 10 criteria.

5.5.4. Final hypotheses

Arriving at this point, we can propose new questions. If the ADNUMERATIVE does not meet the criteria expected for NUMBER or CASE, where do we place the ADNUMERATIVE in the system of features and values? There are two possibilities that I seek to explore in the next section:

1-Treating the ADNUMERATIVE as a feature on its own (H3).

2-Postulating that the ADNUMERATIVE does not comprise a (full) feature in itself, but because it displays properties of other well-defined values (CASE and NUMBER) is a hybrid feature (H4).

Let me address these questions one by one, but I will formulate them as hypotheses:

5.5.4.1. The ADNUMERATIVE as a feature

We have seen that the ADNUMERATIVE does not properly fit into any of the feature values of the language, and of the general typology, overall. As such, is it right to analyse the ADNUMERATIVE as a feature? Let us start by formulating a hypothesis.

H3: The ADNUMERATIVE is a feature.

There are a few arguments to support such a claim. Firstly, the existence of unique or dedicated forms in the paradigm could be used as evidence for postulating a feature (or a value). Secondly, there are syntactic rules that determine it; i.e. it has to be inside of a NumP, but there is a difference between NumPs with lower and higher numerals. Moreover, these syntactic rules have an effect on the morphology; i.e. choosing among the available forms, which seldom can be a unique or dedicated form. Thirdly, inspired by the ideas in Corbett (2012: 262), we can speculate about the existence of a ‘hyper-feature’ or ‘feature-umbrella’ called AGREEMENT, under which NUMBER, CASE and GENDER would appear as the features (or values) derived from it. Thus, ADNUMERATIVE or ‘NUMERACY’ (following the term used in the other analyses) would be a direct morphosyntactic effect of AGREEMENT. The resulting structure looks somehow like Figure 9:

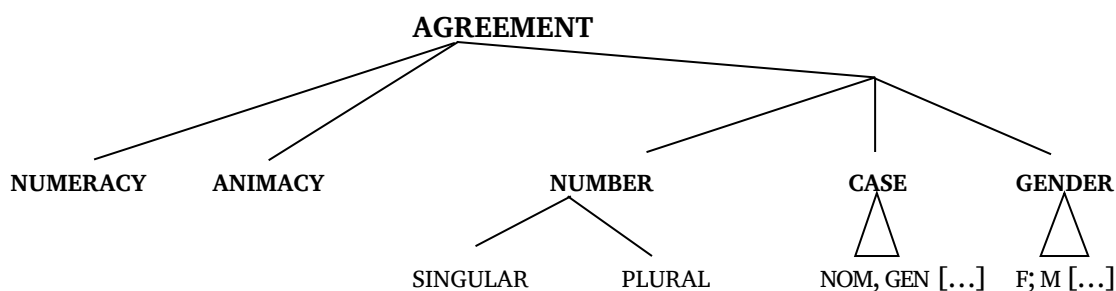


Figure 9 Structure of features and their values predicted by H3.

One of the advantages of this model (or representation) is that it attributes the origin of the ADNUMERATIVE to purely syntactic causes, whereas other FEATURES (such as NUMBER, and less clearly, CASE and GENDER) have a stronger (or better defined) semantic basis.

Such a model predicts that other morphosyntactic phenomena can equally intersect with other cross-linguistically well-known and robust FEATURES, such as GENDER or NUMBER, in the paradigm (and even be on the same level). For example, ANIMACY, which is far more commonly attested than NUMERACY (at least in my West Polesian corpus), also intersects with these well-known features. The AGREEMENT patterns for ANIMACY are affected by, or vary according to NUMBER,¹⁶⁷ GENDER and CASE. Moreover, ANIMACY determines the type of numeral (pronominal, collective or only cardinal) that can be used in a certain NumP (see Footnote 116) and limits the use of the VOCATIVE (see (§3.4.)). And as we have seen in Table 16, NUMERACY and ANIMACY can enter into a conflict which can lead to various possible resolutions.

Corbett (2012), who advocates a simple typology of morphosyntactic features, considers such analyses to be very problematic. Firstly, according to Corbett (2012: 261-262),

¹⁶⁷ Such as NON-HUMAN ANIMATE nouns (see (§3.4.)).

morphosyntactic features tend to be very limited in number. In each language there are always just a handful of them (from which other values can be derived) and they tend to be recurrent from one language to another (i.e. quasi-universal).¹⁶⁸ Thus, creating additional features (instead of extra feature values) creates an “undesirably complex” system for typological work.

Secondly, Corbett (2012) considers that hypothetical features like **AGREEMENT** would be composed of other features (placed almost on the same hierarchical level), which overcomplicate the system, and thus make the work of the typologist a lot harder:¹⁶⁹

“The feature **AGREEMENT** is complex, having as its value other features and values. The point is that the feature **AGREEMENT** is of a different type [...] compared with morphosyntactic features like **GENDER** and **NUMBER**. **NUMBER** is an agreement feature, and can be part of a typology in which agreement features are compared. But if we were to allow similar hierarchical relations *between* the morphosyntactic features, then the space of possibilities would be changed radically” (Corbett 2012: 262).

That is to say, **AGREEMENT** does not (at least, exclusively) belong to morphosyntax (like the rest of the features), because it is a purely syntactic phenomenon. Therefore, there is a risk of comparing different phenomena happening at different levels without realising the gap.

¹⁶⁸ ‘Quasi-universal’ in the sense of ‘generalisation’, or what we expect to find.

¹⁶⁹ Corbett’s idea, and the whole framework of Canonical Typology, seeks to define a “base” (Corbett 2008) (or set) of simple and cross-linguistically recurrent feature values and phenomena, which will allow us to “compar[e] like with like” (Bond 2018: 410).

5.5.4.2. The ADNUMERATIVE as a hybrid feature value

Given that all the previous hypotheses were problematic, let me rephrase a final hypothesis.

H4: The ADNUMERATIVE is a hybrid morphosyntactic feature (a combination of two features), but not a full morphosyntactic feature in itself.

Such a hypothesis predicts the structure in Figure 10:

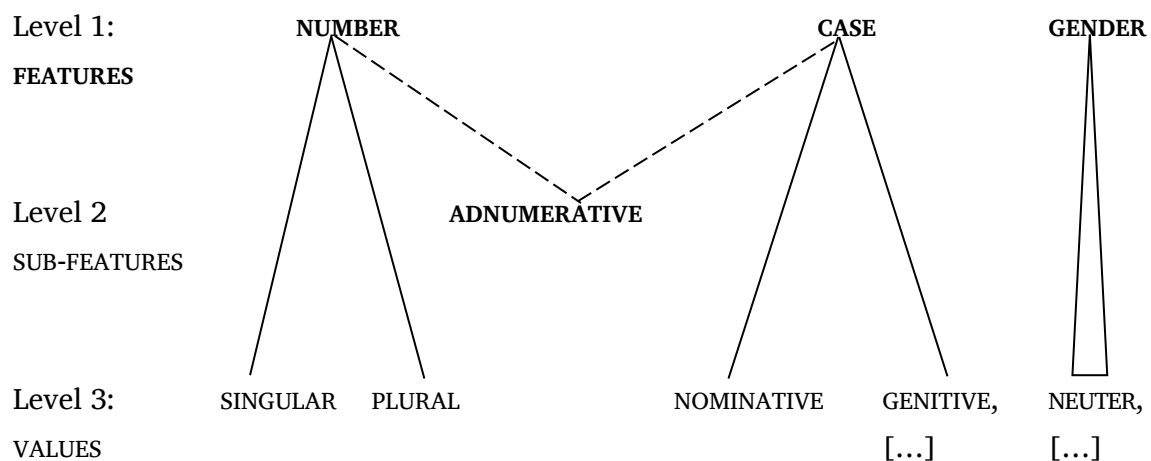


Figure 10 Structure of features and their values predicted by H4.

At first glance, H4 seems a plausible explanation: it allows us to pick the features from **NUMBER** and **CASE** that we like, without fully having to commit to any of them. Yet, taking such an analysis means we are taking an indulgence: creating a sub-feature hybrid level, and thus, the expectation to find similar patterns cross-linguistically. Corbett's (2012) handbook on morphosyntactic features does not give much information about them, but he makes the following remark:

“[...] the penumbra, the less clear area of feature systems, is important for understanding diachrony. It would be hard to imagine how a language could switch from having one clear-cut feature system to another. It is the penumbra of the system which offers potential routes through which features can arise and die out” (Corbett, 2012: 199).

Thus, in a certain way, Corbett (2012) recognises the existence of intermediate levels between features and the importance of studying them. His comment also suggests that they are somehow diachronically unstable, which is in line with Nurmio & Willis (2016) predicted about the fate of the ADNUMERATIVE. This is true for many languages, but BCMS, West Polesian or Bulgarian/Macedonian ADNUMERATIVES, i.a., are counterexamples to such an assumption.¹⁷⁰

The origin of the ADNUMERATIVE can be clearly traced back to the erosion of the DUAL, thus, it has its origin on the NOM/ACC cell of a NUMBER value. With the time, this NUMBER value would have given up some of its canonical properties (such as desemanticising, only becoming available for nouns, becoming syntactically bound) and as a result, it would have started transitioning into CASE. But as canonicity tests show, this is not a clear-cut or canonical CASE, either. I have also speculated about the existence of the GREATER ADNUMERATIVE; and whether it may be an emergent form (morphophonologically) created by analogy with the ADNUMERATIVE. Yet, we have seen that historically nouns headed by higher numerals used to take GENITIVE PLURAL. And synchronically, I have shown evidence from morphology and syntax (in (§5.3.))

¹⁷⁰ Yet, according to Greville G. Corbett (p.c.) numeral phrases are generally in the fringe of the system; thus, they are more likely to retain certain properties over time.

that for certain nouns it is not accurate to call this form a GENITIVE PLURAL, at least a straightforward one.

Perhaps the closest parallel with this in Slavonic languages are the “numeral squishes” (described in great detail in (Corbett 1978b)).¹⁷¹ Corbett’s idea is that most cardinal numerals in Slavonic are not purely adjective-like or noun-like, but that they are on a ‘quasi-continuum’ (i.e. the higher the numeral, the more noun-like). Thus, following the same principle, the ADNUMERATIVE would be on a quasi-continuum of feature values (but without disclosing all the properties of a specific value).

However, the differences are greater than the resemblances. Corbett’s idea of numeral squishes does not deal with features and their values, but with parts of speech within the nominal category. This is an important remark, since the accuracy of traditional classification of parts of speech has been long debated in linguistics (especially in typology) and there is solid evidence supporting such analyses (see (Dixon 1982, Thompson 1989) just to mention a few classic works).¹⁷² Moreover, the taxonomic ambiguity of the numerals described by Corbett (1978b) is for parts of speech within the same domain; i.e. nominals, where it is commonly accepted that there is a great fluidity or ambivalence (even in the context of traditional descriptions of European languages). To bring the analogy back to the domain we are concerned about, it would

¹⁷¹ “A matrix of this type shows that instead of having a set of discrete syntactic categories we are dealing with a quasi-continuum. This phenomenon is what Ross has termed a ‘squish’ (Ross 1972, 1973a)” (Corbett 1978:47).

¹⁷² For example, in his description of Kayardild, Evans (1995) points out that the traditional concept or terms like ‘verb’, ‘noun’ or ‘adjective’ fail to describe what is happening in Kayardild. Instead, he divides the parts of speech into two big groups “nominal words” and “verbal words”.

be like comparing a PLURAL with a GREATER PLURAL and pointing out that the differences between the two are not neat.¹⁷³

Outside the framework of Canonical Typology, Bybee (1985) gives a description of the ANTERIOR as an intermediate or hybrid feature value between TENSE and ASPECT:

“It seems to resemble a tense more than an aspect, since it does not affect the internal temporal contours of the situation. Since *anterior* deals with time, it does not seem to be a mood. [...] This would suggest that it is a category which is independent of the present and past” (Bybee 1985: 160).

Nonetheless, Corbett’s (2013) later work looks suspiciously at intermediate feature values:

“This weakening would be more serious; if we allowed values to belong to more than one feature, an inventory of possible features and their values would become an elusive goal. While our suggested example may seem unlikely, there are published analyses which treat particular values as belonging to an unexpected feature, thus allowing them to belong to two features. Given the implications of this move, we should not allow this type of weakening if at all possible. I believe that the suggested analyses of this type are not optimal: there are better analyses for the data and these follow the requirement that a value should belong to just one feature [...]” (Corbett 2013: 53).

¹⁷³ But not the difference between CASE and NUMBER, which operate at higher levels.

Although, in favour of this hypothesis, it seems that Corbett's concern has to do with well-established feature values (e.g. PLURAL) being used as a value of another value (e.g. GENDER), whilst claiming that they retain properties of both. When it comes to the ADNUMERATIVE, we have seen that there is little known or written about it, in part, because it is a marginal phenomenon cross-linguistically. Thus, there is little or no consensus about how to analyse it (and as we have seen, different positions are found in the literature). Moreover, H4 does not claim that the ADNUMERATIVE is *simultaneously* a CASE and a NUMBER value. Rather, that the ADNUMERATIVE is somewhere between both, and so Corbett's objection does not necessarily cover instances like this.

5.6. Summary and conclusions

I have started by presenting the adnumerative and proposing two possible analyses or hypotheses to deal with them:

H1: West Polesian adnumerative is a NUMBER value.

H2: West Polesian adnumerative is a CASE value.

I have given a cross-linguistic overview of adnumerative forms, although as noted earlier, I have only found instances of adnumerative forms documented in Indo-European languages (§5.1.2.). All had in common that the forms are relatively infrequent (and thus, seemingly fossilised), usually limited to a few nouns or a specific noun class; syntactically non-autonomous; only available for nouns, and only in DIRECT CASES. Moreover, the origin of the ADNUMERATIVE was traceable to an erosion or shift of a NUMBER distinction.

I have shown how the ADNUMERATIVE is morphologically very robust and marked consistently in every countable noun (§5.2.2.); whereas the alleged GREATER ADNUMERATIVE is harder to postulate as a separate form based entirely on morphophonology (as it would be too often syncretic with GEN PL) (§5.3.5.). The best analysis of these special ‘GREATER ADNUMERATIVE’ forms has been to treat them as a sub-CASE value of the GENITIVE PLURAL, which has emerged thanks to the overabundance of forms in that cell. I have also shown that morphologically the ADNUMERATIVE (or the WP adnumerative forms) depends on the PLURAL sub-paradigm for its NON-DIRECT CASES. I have explained some of the peculiarities of agreement of NumPs containing nouns in these two forms with verbs and other nominals (§5.2.3.). I have shown that in some varieties, NumPs with adnumerative(s) can even create ‘a shield’ that overrides the effects of ANIMACY (§5.2.3.1.1.).

I then moved to a level of typological analysis (§5.5.). I have started by arguing in favour of treating the ADNUMERATIVE as a NUMBER value (H1) (§5.5.1.). The strongest arguments in favour of this have been the etymology of the ADNUMERATIVE, its close relation to quantification/individuation (semantically), and the fact that it is in complementary distribution with DIRECT CASES. I have presented a detailed description of which criteria determine a canonical NUMBER value (§5.5.1.1.), and how the ADNUMERATIVE fails to meet some of them (§5.5.1.2.).

I have postulated that the ADNUMERATIVE could be a CASE value (H2) (§5.5.2.). The strongest argument for this has been that it marks a syntactic relation between a head (numeral) and a target (noun). Yet, the fact that it has to fulfil the role of two CASES at the same time (NOM or ACC + ADNUM) has been one of the major objections to this hypothesis.

I have reconsidered (H1) and (H2) and looked at the **ADNUMERATIVE** in the light of the standard criteria applied by Canonical Typology for canonical features and their values, and we have seen that the **ADNUMERATIVE** falls short of the criteria regardless of the analysis (§5.5.3.). This has taken me to propose a third hypothesis: since the **ADNUMERATIVE** has a dedicated form, that the **ADNUMERATIVE** would be a **FEATURE** on its own (H3) (§5.5.4.1.). However, the **ADNUMERATIVE** does not look like what we would expect for a **FEATURE**. Firstly, we have said that features tend to be very scanty and quasi-universal typologically; whereas the **ADNUMERATIVE** has only been documented in a handful of Indo-European languages, so far, and it is a marginal phenomenon inside the language systems themselves. Secondly, features depend on morphosyntax and semantics, whereas the **ADNUMERATIVE** would rely solely on syntax. Thirdly, the **ADNUMERATIVE** greatly resembles the behaviour of the values (of the features) in that it cannot cross-cut other features on its own; it always appears where we would expect a **NUMBER** or **CASE**.¹⁷⁴ When it comes to H1 and H2, it is true that the behaviour of the **ADNUMERATIVE** was closer to what can be observed from other feature values. Nevertheless, we have seen that the **ADNUMERATIVE** fails to fulfil the criteria of both canonical **NUMBER** and **CASE** (§5.5.3.).

This has led me to finally propose that the **ADNUMERATIVE** could be a feature value between **CASE** and **NUMBER** (H4) (§5.5.4.2.). However, given the limitations of the system of morphosyntactic features to describe phenomena like this, I can only discern that the **ADNUMERATIVE** is mapped somewhere in the penumbra between the values for **CASE** and **NUMBER**. Hence, I conclude that H4 seems to be the better explanation,

¹⁷⁴ Except for **ANIMACY**, in some contexts, but if we analyse it as a sub-gender value (following Corbett's (1991) approach, then no feature would be directly intersected by **NUMERACY/ADNUMERATIVE** on its own.

admitting that there is some vagueness in the explanation. At the same time, this points out an important discovery. We would expect that feature values which are between two features to be diachronically unstable (i.e. either destined to extinction or to a ‘full transition’ into a single **FEATURE** value). However, we have seen that West Polesian, Bulgarian, Macedonian and BCMS adnumeratives (i.a.) can remain in that non-binary or ambiguous position for centuries.

The study of the **ADNUMERATIVE** in West Polesian opens certain questions for further research:

- a) How frequent are “hybrid features” and how or where do we map them into the system of morphosyntactic features and their values? Should we reform the system for describing them?¹⁷⁵
- b) West Polesian, together with BCMS, probably has the most (morphologically) robust **ADNUMERATIVE** form ever documented. Can we apply any piece of knowledge extracted from West Polesian to other **ADNUMERATIVE** forms in other languages, or is **ADNUMERATIVE** simply a very vague term?
- c) Should the analyses proposed for Eastern Slavonic numeral phrases be revised in the data of West Polesian? Could all the apparent **GENITIVE** forms (in Russian) be allomorphs of **ADNUMERATIVE** forms? Do numerals govern or agree with the noun, given all the possible configurations of numeral + adjective + noun?

¹⁷⁵ One of the main problems for addressing the question of features or values in the penumbra is that often there is not a clear consensus on what exactly features are. For example, we have seen that many authors considered the **ADNUMERATIVE** to be a **CASE** or a **NUMBER** value, but they do not provide a definition of what they understand by a **CASE** or a **NUMBER** value.

Chapter 6

Canonical suppletion in West Polesian:

the nouns ‘year’ and ‘person’

Suppletion is a lexical split that happens when a lexeme displays different roots or stems in its paradigm under different conditions, as in English *good* > *bett-er* (**good-er*; vs. *smart* > *smart-er*); (Bybee 1985, Chumakina et al. 2004, Corbett 2007, Mel’čuk 1994). For this chapter, I am going to focus on two suppletive nouns in West Polesian: ‘year’ and ‘person’. At first glance, the study of these nouns may not seem particularly innovative, as they are suppletive in Russian and other Slavonic languages as well. However, data from West Polesian reveal a particular distribution of suppletive stems and a much higher level of complexity, given that multiple stems are involved and more combinations are possible. I take a “canonical approach” (CT) to analyse the nouns ‘year’ and ‘person’. And since I am going to study two different nouns, different results can already be predicted for each of them. More specifically, I show that the noun ‘year’ in West Polesian displays a behaviour that approaches in many aspects the “canonical instance” of suppletion (Corbett 2007) (from a typological perspective); i.e. the most unpredictable or irregular suppletive paradigm ever possible.¹⁷⁶

¹⁷⁶ “Canonical instances are those that match the canon: they are the best, the clearest, the indisputable ones” (Corbett 2010: 141).

I start by introducing suppletion and the problem of the nouns ‘year’ and ‘person’ in West Polesian (§6.1.). I ask whether all the different forms obtained are overlapping synonyms in reality, or else, how do we explain all the apparent ‘mess’ in the paradigms. Secondly, I make a cross-Slavonic survey of the nouns ‘year’ (§6.2.1.) and ‘person’ (§6.3.1.). I contrast the results of the survey with data from West Polesian and explain the possible origin of the different stems (§6.2.2.), and hopefully use it as evidence for stating that all these stems are not synonyms. Thirdly, I propose some general patterns for suppletion for both nouns (§6.2.3.; §6.3.2.). I show how West Polesian speakers can combine several suppletive stems under different conditions with no apparent semantic or sociolinguistic correlate (§6.2.4.; §6.3.3.). This will provide evidence to challenge some traditional assumptions of suppletion and overabundance (§6.4.), which will call for a re-examination. After this, in (§6.5.) I look at these phenomena in the light of Canonical Typology (CT). I examine how the nouns ‘person’ and particularly ‘year’ approach the most extreme possible instances of “canonical suppletion”. Finally, in (§6.6.) I present a summary of the results and draw general conclusions. To sum up, I explore how these two specific lexical items in West Polesian can enrich our knowledge of morphological typology in the areas of suppletion and overabundance.

6.1. Introduction

According to Bobaljik (2012: 27), the term suppletion was first coined by Osthoff (1888, 1899). Suppletion (understood in its narrowest sense)¹⁷⁷ is a split in an inflectional paradigm, which causes certain cells to use phonologically distinct stems, and which cannot be synchronically derived by morphophonological rules. I give more details on suppletion later on, but as an example we can think of the English verb *to go* in the PAST SIMPLE *went* (**goed*; as in *jump* > *jumped*). Both stems are phonologically very different; indeed, they have different etymologies (Oxford English Dictionary 2018). Hence, *go* and *went* hold a suppletive relation; that is to say, their relation is semantic, rather than formal (phonological). There are few suppletive nouns in West Polesian (or at least with some partial stem alternation), but the nouns ‘year’ and ‘person’ present an extreme case, as I will be showing throughout this chapter. For this matter, first, (§6.1.1.) I present the problem of suppletion in West Polesian, in comparison to what readers may know about Russian. Second, (§6.1.2.) I give some remarks about the origin of the data in West Polesian for this chapter. And third, (§6.1.3.) since the distribution of suppletion for these nouns is tightly linked to quantification, I give a brief review of numeral phrases in Slavonic languages.

¹⁷⁷ I am aware that other authors, such as Mel’čuk (1994), have more inclusive definitions of suppletion (e.g. covering derivation and inflection; or even inclusive of PHRASAL SUPPLETION). I need to make clear that, for the sake of clarity, I only deal with suppletion in inflectional paradigms in this work. This is also in line with Corbett’s (2007) typology, which I use as a foundation for my research on suppletion.

6.1.1. Presentation of the problem

At first glance, a reader who is familiar with Russian grammar may not be particularly impressed by the fact that the nouns ‘year’ and ‘person’ are suppletive, as they are also suppletive in Russian (106) and (107).

Russian

- (106) a. Ja ži-l (odin) **god** v Moskv-e.
1SG.NOM live-PST.SG.M one.NOM.SG year.NOM.SG in Moscow-LOC.SG

‘I lived for a year in Moscow.’

- b.^I Ja ži-l pjat’ **let** v Moskv-e.
1SG.NOM live-PST.SG.M five.NOM year.GEN.PL in Moscow-LOC.SG

- b.^{II} *Ja ži-l pjat’ **god-ov** v Moskv-e.
1SG.NOM live-PST.SG.M five.NOM year-GEN.PL in Moscow-LOC.SG

‘I lived for five years in Moscow.’

Russian

- (107) a. Poprosi-l den’g-i odn-omu **čelovek-u.**
ask.PRF-PST.M.SG money-ACC.PL one-DAT.SG.M person(M)-DAT.SG

‘[He] asked one person for money.’

- b.^I Poprosi-l den’g-i mnog-im **ljudj-am**
ask.PRF-PST.M.SG money-ACC.PL many-DAT.PL person(M)-DAT.PL

- b.^{II} *Poprosil den’g-i mnog-im **čelovek-am.**
ask.PRF-PST.M.SG money-ACC.PL many-DAT.PL person(M)-DAT.PL

‘[He] asked many people for money.’

If we look at data from West Polesian (108), we can see some speakers adjusting to the same pattern, i.e. using suppletive stems that appear in mutually exclusive contexts:

- (108) a. (Z7.1 00:09) [...]
 ni bu-l-o nijak-o(j) zizn-i 'ljudj-am (*tʃolovik-am)[...].
 NEG be-PST-N.SG any-GEN.SG life-GEN.SG people-DAT.PL
 'It was hard for people to live [lit. there was no life for people].'
- b. (Z7.1 00:41) [...]
 da-v fet-ofio mɨk-a, tak-omu nɨdaljok-omu
 give.PRF-PST that-GEN.SG bag-GEN.SG that-DAT.SG fool-DAT.SG
 tʃolo'vik-u (*ljudj-u).
 person-DAT.SG
 '[He] gave the bag to a not very smart man/person.'

But there is a better story to be told. Let us look at the three utterances in (109) produced by the speaker TL6. Note that they were produced within a span of less than fifteen minutes, in the same context (same room, same hearers and same style).

- (109) a. (TL6.5 00:45)
 uʒe sim ro'kru v jomu bu-l-o
 already seven year-GEN.PL/GRADNM in 3SG.DAT.M be-PST-N.SG
 'He was already seven years old.'
- b. (TL6.5 02:40)
 v sorok lit vin i umer
 in forty year-GEN.PL/GRADNM 3SG.NOM.M and die.PST.M.SG
 'He died when he was forty years old.'

[But some minutes later]

(TL6.7 02:00)

- c. oj, i **mnɪfɪa** uʒe minji **fɪa'd-ov** uʒe,
oh and many already 1SG.DAT year-GEN.PL already
- dɪvɪnosta** **dva** 'roki!
 ninety two.NOM.M year.ADNM

‘Oh, I’m already very old (lit. many of years old); ninety-two years.’

We can observe that the speaker TL6 used three different stems for the noun ‘year’ for what should be the same cell of the paradigm (GENITIVE PLURAL / GREATER ADNUMERATIVE (GRADNM)): *sim ro'kiʋ* ‘seven years’; *sorok lit* ‘forty years’; *mnɪfɪa fɪa'dov* ‘many years’. How can such a complex paradigm be possible? Are not all these stems just mere synonyms? I will try to address this issue in the following sections.

6.1.2. Methodological remarks

For the premises of this chapter, I concentrate on two specific lexical items (the noun ‘year’ and the noun ‘person’), yet I seek the bigger typological picture; that is to say, what do these nouns teach us about suppletion.

The data for this study have been extracted from the same corpus of recordings I have already introduced (Chapter 2), which records age, gender, diatopic and idiolectal variation. Yet, it must also be said that there is hardly any class or style distinction in the corpus. Speaking West Polesian is a sign of belonging to a rural area (and thus to an economically, but especially socially disfavoured class) and it is used only for informal

oral communication. The subsection of the corpus dealing with the nouns ‘year’ and ‘person’ has been exclusively composed of utterances from free texts (i.e. they have not been directly elicited) and a couple of overheard sentences. This way, I have gathered a total of 401 tokens of the noun ‘year’, and 245 for ‘person’. The amount of tokens for the noun ‘person’ is significantly smaller than for the noun ‘year’ in the corpus. As a result, any conclusions drawn from the data for ‘person’ are less robust than for ‘year’. According to Greville G. Corbett (p.c.), this imbalance might be related to the fact that the noun ‘person’ is in competition with other nouns (such as ‘woman’ or ‘boy’); whereas the noun ‘year’ is very frequent but hardly replaceable by any synonym or paraphrase.

6.1.3. Remarks about Slavonic counting systems and suppletion

Before we go further in this chapter, it is important to remark that in West Polesian, as in other Slavonic varieties, most of the suppletive stem alternation happens in the cells used in the presence of numerals or quantifiers (adnumerative forms, which can be morphophonologically dedicated or not). Thus, it is crucial to have an understanding of the functions of these cells in Slavonic languages in order to understand the phenomenon. For this matter see more on adnumeratives in the previous chapter (Chapter 5); as well as morphosyntactic phenomena around numerals in (Chapters 4, 5). This is the reason why I have included this chapter under the section on quantification (*Section I.*) and why I concentrate on the interactions between suppletive stems and quantifiers for the purposes of this chapter.¹⁷⁸

¹⁷⁸ I have already discussed the potential GREATER ADNUMERATIVE form in (§5.3) and concluded that, so far, it was better described as an alternative SECOND GENITIVE (PLURAL). However, in order not to confuse it with the SECOND GENITIVE (SINGULAR) (§3.2.1.), which has a very different function, I prefer to keep

6.2. The noun ‘year’.

In some Slavonic languages, the noun ‘year’ also displays suppletion, mostly conditioned by the type of numeral phrase governing the NP (i.e. in Russian [one] *god*; [five] *let*). What is interesting about West Polesian dialects is that they concentrate *all* the lexical roots and the suppletion patterns found across Slavonic varieties (*fi**od*, *let*, *ri**k*), with some additional innovations. That is to say, whilst in the rest of Slavonic languages it is a two-way distinction; in West Polesian it is a three-way one, making it a lot more complex and opening the door to more possible combinations (as I show below, particularly in Table 30). Likewise, very often even a single idiolect (i.e. a single person’s own variety) can concentrate multiple patterns and stems (as in (109)). As a result, two “non-canonical” (Corbett 2005, 2007) phenomena in typology come together in the noun ‘year’: suppletion and overabundance; i.e. “more than one form in a cell” (Thornton, 2011). What is more, as I show in (§6.5.), this particular lexical item is close to the “canonical instance of suppletion”. In Corbett’s (2007) understanding “canonical suppletion” means the most unpredictable paradigm; the most phonologically distinct stems; but the most semantically regular correlation between them. Thus, after presenting (§6.2.1.) the incidence of the suppletive stems for ‘year’ in the Slavonic and their etymology (§6.2.2.); I compare this to the complex results found in the corpus of West Polesian (§6.2.3.); and try to set some conditions which could be behind the choice of the different suppletive stems (§6.2.4.).

the name GREATER ADNUMERATIVE. In any case, for the sake of consistency in the paradigms illustrated here, I represent the GREATER ADNUMERATIVE as closely related to the (LOWER) ADNUMERATIVE, when it is morphophonologically different from the regular GENITIVE PLURAL. Otherwise, I treat all the noun forms headed by higher numerals as GENITIVE PLURAL.

6.2.1. The noun ‘year’ across the Slavonic family

As has been already said, the noun ‘year’ has a suppletive paradigm in many Slavonic languages. Thus, visualising the suppletive stems and the available distributions for these across the Slavonic family can bring light to the study. For that reason, I have made a survey of the noun ‘year’ in several languages from each of the three main Slavonic subfamilies or subgroups. The materials for the survey are based on *ParaSol, a Corpus of Slavic and Other Languages* (Waldenfels & Meyer 2011). As I said in (§6.1.3.), the cells most affected by suppletion are the ones that are most often employed for numerals (i.e. ADNUMERATIVE (ADNM), GENITIVE PLURAL/GREATER ADNUMERATIVE (GRADNM) or similar). That is why, in the following tables (Table 26, Table 27), I pay special attention to the type of numeral each form appears with in DIRECT CASES.

Here is a list of the languages included in the survey, classified according to their subfamilies:

Southern Slavonic: Bulgarian (BG); Croatian (HR)¹⁷⁹; Slovene (SVO).

Western Slavonic: Czech (CZ); Polish (POL); Slovak (SK); Upper Sorbian (US).

Eastern Slavonic: (Standard) Belarusian (BLM); Contemporary Standard Russian (CSR); (Standard) Ukrainian (ULM).

¹⁷⁹ In the survey Macedonian (MKD) provided the same results as Bulgarian, and Serbian (SRB) provided the same results as Croatian (HR). For this reason I only present one member of each pair.

Because of the typological orientation of the present work, I have decided to group the data from languages according to their suppletion patterns, rather than their areal/genetic affiliation (Table 26, Table 27).

Table 26 Languages with no suppletive paradigms for ‘year’

	NOM SG	LOWER NUMERALS		HIGHER NUMERALS
BLM	hod	hody		hadoů
BG ¹⁸⁰	godina	godini		godini
HR	godina	godine		godina
SK	rok	roky ¹⁸¹		rokov
ULM	rik	roky		rokiv
SVO	leto	[2] leti	[3-4] leta	let
US	lěto	[2] lěće	[3-4] lěta/lět ¹⁸²	lět

¹⁸⁰ The lower/higher numeral distinction is not relevant at all for Bulgarian or Macedonian, yet for the sake of consistency, I applied the same questions for them.

¹⁸¹ But ‘twenty two’ appears with *rokov* (e.g. it is treated like a higher numeral, rather than *20 + 2 roky* pattern, like Polish, i.a.), and thus, I assume all comparably structured numerals behave according to this pattern.

¹⁸² I was unable to find any results for this noun in the *ParaSol* corpus (Waldenfels & Meyer 2011). However, Schuster-Šewc (1999) points out that even though the standard (or ‘literary norm’) is *štyri leta* ‘four years’, nowadays western dialects often use *štyri lět*, instead. In any case, the choice of inflectional suffixes does not alter the form of the stem, which is our main concern in this chapter.

Table 27 Languages with more than one stem for ‘year’

	NOM SG	LOWER NUMERALS	HIGHER NUMERALS
CZ	rok	roky	let
POL	rok	lata	lat
CSR	god	goda	let

The examples from the Southern Slavonic subfamily do not reveal anything interesting in terms of suppletion. However, I have decided to include them in order to show how their stems for ‘year’ and non-suppletive paradigms are also found in West Polesian.

6.2.2. Etymology

According to Chumakina et al. (2004: 289-290), originally Common Eastern Slavonic (CES) used the form **lěto-*, which in its primary sense meant ‘summer’ and is already documented in twelfth-century writings.¹⁸³ Later on the form *god-* (which originally meant ‘period’ or ‘time of a day’) appeared as an intrusion and has been slowly taking over the whole paradigm for centuries, to the point of leaving the form *let-* just available for the GENITIVE PLURAL.

But that only explains part of the stems. Western and Eastern Slavonic (Standard Ukrainian, and as I have already shown in (109) a., West Polesian as well) varieties

¹⁸³ According to Moszyński (2006: 199), the Common Slavonic form could be either **lěto* or **lāto*.

show a different lexical stem: *rok-*. According to Boldyrjev et al. (2006) and Vasmer (1965), the form *rok-* derives from the Proto-Slavonic **rokъ-* which meant ‘period, term, time’ (which shares a common origin with the Contemporary Standard Russian (CSR) *srok*, with the same meaning).¹⁸⁴ Now, since the three stems existed in Proto-Slavonic and they spread heterogeneously among different Slavonic languages it seems difficult to determine whether **lět-* or **rokъ-* is the original form for ‘year’. Brückner’s (1927) comments for Polish suggest that **lět-* was the original form. Thus, the form **rokъ-* would have fully taken over Ukrainian and Slovak paradigms; whereas only partially in Polish and Czech, resulting in a pattern of suppletion which resembles the one in Russian. Moreover, this means that Slovene and Upper Sorbian are the most conservative Slavonic varieties in this respect.

6.2.3. West Polesian suppletion patterns for ‘year’

Based on what we know about most nouns in West Polesian and the different forms of the noun ‘year’ recorded in the corpus, we can reconstruct the inflectional paradigms in Table 28 a-c.

¹⁸⁴ According to Vasmer (1965), the Proto-Slavonic form passed into Common Eastern Slavonic without alterations (i.e. as *rokъ*).

Table 28 Reconstruction of inflectional paradigms of the noun ‘year’

Table 28 a.			Table 28 b.			Table 28 c. ¹⁸⁵		
	SINGULAR	PLURAL		SINGULAR	PLURAL		SINGULAR	PLURAL
NOM	fiɔd	fiɔɫ	NOM	riɫ	roɫ	NOM		
ACC	fiɔd	fiɔɫ	ACC	riɫ	roɫ	ACC		
GEN	fiɔd-a -u	fiɔɔv	GEN	roku	roɫv	GEN		
DAT	fiɔɔ	fiɔam	DAT	roku	roɫam	DAT		
INST	fiɔɔm	fiɔamɪ	INST	roɫm	roɫamɪ	INST		
LOC	fiɔɔ	fiɔax	LOC	roku	roɫax	LOC		litax
ADNM	fiɔɫ	fiɔd	ADNM	roɫ	riɫ	ADNM		lit

As a possible result of language contact, West Polesian has adopted three different stems (with their respective paradigms) for the noun ‘year’.¹⁸⁶ We would expect the speakers to stick to one of these paradigms, and some do (especially to Table 28 a.). However, in the case of Table 28 c., a combination with a different pattern is needed to fill in the blanks. Note that the lexeme *lit-* also exists independently (with a non-defective paradigm), although synchronically it only means ‘summer’.

But as I have already shown, the patterns shown above tend to be rather an ideal. In Table 29 there are some of the most robust examples of different speakers, from

¹⁸⁵ NB. In Table 28 c. the cells that have not been attested as a form for ‘year’ have been shaded in grey.

¹⁸⁶ It is worth pointing out the Ukrainian-style vocalism of Table 28 b-c. (realisation of stressed /o/ and ɔ (i.e. /ě/) as /i/), which may confuse the reader. According to Matthew Baerman (p.c.), this provides clear evidence that once-distinct varieties are being mixed, from a Slavic perspective. See more on phonology in (§3.1).

different villages (represented by the letters of their code), based on utterances from texts. This table shows how much variation is possible, especially when it comes to the GEN PL / GRADNM cell, causing overabundance. Note that there is great variation between the inflections (i.e. suffixes) each speaker assigns to each form of the cell, especially in those cells that are somehow related to adnumeratives and quantification (GEN SG, NOM PL, ADNUMERATIVE and GEN PL /GRADNM). That is a topic for research on its own. Moreover, there is variation in the phonological realisation of some of the stems; e.g. *lit* - *ljet*; *rik* - *rok*-. However, for the purposes of this chapter, I only focus on the existing variation between stems for ‘year’. In order to help the reader to identify them easily and not get distracted by the aforementioned variation, I present the cells containing those stems in the form of a colour code: *hod*-, blue; *rik*-, red; *lit*-, yellow.

It is interesting to point out that there is a high level of variation within the same variety. HL2 (female) and HL3 (male) being husband and wife display different patterns, whereas HL4’s (male) matches HL2’s.

Table 29 Distribution of suppletive stems according to governing numerals

	NOM SG	LOWER NUMERALS	HIGHER NUMERALS
B6	fod	'fodi	fod
T3	fod	'foda	lit
HL4	fod	'foda	lit
HL2		'foda	ljet
Tor1	fod	'foda	rik
			'rokiu
			fod
			fio'dov
Xab1		fio'da	ro'kiu
			lit
TL6		'roki	ro'kiu
		'foda	lit
HL3	rik	'roki	lit
		fio'da	fio'dov

6.2.4. Conditions. Are there any rules for the distribution of the stems?

As already mentioned, some speakers adhere more consistently to a pattern, whereas other speakers constantly alternate between them for apparently no particular reason.

Age does not seem to be a conditioning factor,¹⁸⁷ nor style, as the texts recorded addressed similar or the same topics, and most speakers did not use West Polesian (at least purely) outside their family and village context (with people they know). Moreover, *all* the forms mentioned here were recorded in free texts and conversations between the speakers (i.e. not elicited directly, which could have affected the reliability of the answers). This seems to be a quite canonical case of overabundance (Thornton 2011, 2013, 2019), that is to say, more than one form is available for the same cell (or function).

Before stating that there is pure overabundance, which has traditionally been seen as a result of an inaccurate or superficial analysis by many linguists, I have decided to test the stems in different syntactic contexts or conditions to see if they restrict the use of certain stems. I have taken some ideas for such conditions from Bortnik (1979) (in Chumakina et al. (2004: 294), see references there) which also influence the suppletive patterns in Russian for its cognates ‘year’ and ‘person’. Moreover, I have also created additional putative conditions based on observations of the behaviour of numeral phrases in the entire corpus. That is to say, far beyond the group of suppletive nouns, different constructions (including quantifiers, various numerals or adverbs) have different effects on NPs and agreement in West Polesian.

¹⁸⁷ Note that age tends to be closely related to level of education and exposure to Russian. Younger people had more schooling than older people (particularly women), which from 1945 on was primarily conducted in Russian (and less frequently in BLM). Moreover, younger people tend to spend more time in the cities and towns where Russian is the primary language. Hence, the younger the speaker, the most likely to be bilingual with Russian, and hence more likely to code-switch. I give more details on this in (§1.2.) and (§2.3.3.).

6.2.4.1. Unbound or non-governed instances of ‘year’

Amongst the possible instances in which ‘year’ can appear unbound I have listed three: (§6.2.4.1.1.) with no numeral or quantifier at all; (§6.2.4.1.2.) where the noun ‘year’ is used with nominalized forms of the numerals; and (§6.2.4.1.3.) with ordinal numerals.¹⁸⁸

6.2.4.1.1. Unbound or non-governed ‘year’

In Russian only the cognate stem *god-* is allowed to appear ungoverned (in a noun), as the ungoverned stem *let-* denotes ‘summer’.¹⁸⁹ In West Polesian, when it is not governed by a numeral the *most common stem* by far is *fiod-*, even though *rik-* is also possible. The only stem that cannot appear unbound is *lit-* for all the people interviewed (B8 being the only exception so far).

(110) a. (B8.5 02:46)

uʒe povirova-u, uʒe v star-ix le't-ax.
already believe.PRF-PST.M.SG already in old-LOC.PL year-LOC.PL
‘[He] became a believer already in his late (lit. ‘old’) years.’

b. (TL1.1 06:58)

slava Bofi-u, priʃ-l-i i fiet-i-je fio'da
glory.NOM? God-DAT.SG arrive.PRF-PST-PL and this-NOM.PL year.NOM.PL
‘Thank God, these years arrived.’

¹⁸⁸ Chumakina et al. (2004: 294) distinguish between “adjective + noun” and “ordinal + noun” in their conditions affecting the cognate form in Russian. However, such conditions do not give the different results in West Polesian.

¹⁸⁹ Although *let-* is permitted if it is following a possessor, according to Bortnik (1979).

c. (Z7.4.2 03:24) [...]

u njak-u por-u 'roku
in any-LOC.SG season-LOC.SG year-GEN.SG

‘[...] in any time (lit. season) of the year.’

d. (B19.1P 00:01)

bifi dava-u 'fodi ljudj-am i živjol-am
God.NOM.SG give-PST.M.SG year.NOM.PL person-DAT.PL and animal-DAT.PL

‘God gave years to people and animals.’

6.2.4.1.2. Status numerals

Status numerals (STAT) are those highly lexicalised and/or frequent forms which refer to specific age or time periods (e.g. ‘the eighties’). The corpus is small and so the fact that I have not been able to record certain instances does not necessarily imply that they do not exist. In spite of the many restrictions for suppletive stems, in Russian these type of constructions require the stem *let-* in the GEN PL cell and *god-* elsewhere (Bortnik 1979). When it comes to status numerals in West Polesian, probably the most natural expression is STATUS NUMERAL + LOC PL. Conversely, this only works for *fiod-* (as in *fio'dax*) and is productive neither for *rk-* nor *lit-* (* *v fįftdisjatix litax*; * *v fįftdisjatix rokax* ‘in the sixties’), and there are no recorded instances of them being used in GEN PL (as in Russian). In any case, status numerals are very rare in the corpus; often they appear without the noun ‘year’ (just an ordinal), and it may be that West Polesian has borrowed the model from Russian (as it is morphologically identical).

(111) a. (Z.4.1.2 03:13)

u fįsdisjat-ix fio'dax, uže potſi-l-i vozi-ti xlib
in sixties-LOC.PL year-LOC.PL already start-PST-PL bring-INF bread.ACC.SG

‘In the sixties, they already started to bring bread.’

b. (HL2.23 00:01) [HL4]

banja[*ov*] sta-l-a v nas v **jak-ix** **fio'd-ax?**
 banja-[DAT.SG] become-PST-F.SG in 1PL.GEN in which-LOC.PL year-LOC.PL

mo **pɨdisjat-ɨj**, tʃɨ ʃɨ**disjat-ɨj**
 maybe fiftieth-NOM.SG.M or sixtieth-NOM.SG.M

‘In which years did they start to run the banja? Maybe in the fifties, or
 sixties.’

6.2.4.1.3. Ordinal numeral + ‘year’

In Russian we only find the stem *god-* after an ordinal numeral; whereas in West Polesian both forms *fiod-* and *rik-* are possible. However, based on observations from the corpus, the form *fiod-* is more common in such syntactic contexts; especially when referring to a specific date, rather than someone's age.

(112) (T3.10 01:26)

uze **dvatsat** ʃost-ɨ **rik**, dvatsat pjet,
 already twenty sixth-NOM.SG.M year.NOM.SG twenty five

dvatsat pajʃ-ov ʃost-ɨ...
 twenty past-PST.M.SG sixth-NOM.SG.M

‘Already the twenty sixth year, twenty five, the twenty sixth had
 started...’

6.2.4.2. Bound or governed forms of ‘year’

Amongst the putative conditions which may affect the distribution of the stems I consider the effects of ‘normal’ cardinal numerals (§6.2.4.2.1.); other quantifiers and Q-words (§6.2.4.2.2.); postnominal cardinal numerals (§6.2.4.2.3.); and approximate

quantities (§6.2.4.2.4.). In addition to this, I examine whether the use of temporal adverbs (§6.2.4.2.5.); perception of the quantity (§6.2.4.2.6.); or ANIMACY (§6.2.4.2.7.) have any effect on the distribution of stems.

6.2.4.2.1. Forms of ‘year’ bound by a cardinal numeral

As I have shown in Table 29, when a cardinal numeral is heading the noun ‘year’ there are many suppletion patterns available, as speakers can combine the three stems quite freely. We have seen that in examples (109), all produced by TL6 within a few minutes; but there are also other examples of different stems used within seconds by a single speaker in the corpus, such as HL2 in (113).

(113) (HL2.4 00:15) [HL3]

tam na Moskv-u puŝ-ov i tam probu-u
there in Moscow-ACC.SG go.PRF-PST.M.SG and there spend-PST.M.SG

tri 'roki, bojova-u. daže bilŝ, usjaŝo vmistje
three.ACC year.ADNM fight-PST.M.SG even more all together

pjetj ŝo'd-ov, bojo'va-l-i
five year-GEN.PL fight-PST-PL

‘[A German] went to Moscow and spent there three years, fighting.

Even more, in total [they] were fighting for five years.’

6.2.4.2.2. With question words (Q) and quantifiers

Apart from the fact that quantifiers (i.e. ‘many’; ‘few’) and interrogatives like ‘how much’ govern GEN PL/ GRADNM, data from the corpus reveal no restrictions for the choice of stem.

(114) a. (Z7.4.2 04:03)

ja uže za kiljki 'rok-iv provri-l-a!
 1SG.NOM already for how_many year-GEN.PL believe.PRF-PST-F.SG

‘After many years I have [finally] believed [it]!’

b. (B4.1 04:23) [B2]

a slux-ajte, a ski[l]ko vaŝ batjk-o
 so listen-IMP.2SG so how_much POSS.2PL.NOM.SG.M father-NOM.SG

prozi-u fiod?
 live.PRF-PST.M.SG year.GRADNM

‘Listen, how many years did your father live?’

c. (T7.7 00:03)

fiet-o 3 skolko lit uže priŝ-l-o,
 that-NOM.SG.N PART how_many year-GEN.PL already arrive.PRF-PST-N.SG

fiet-o sorok lit priŝ-l-o
 that-NOM.SG.N forty year-GEN.PL arrive.PRF-PST-N.SG

‘How many years since that happened, forty years have passed.’

6.2.4.2.3. Before the numeral

As with every other noun in West Polesian, where the noun precedes the numeral, the noun takes GEN PL / GRADNM even with lower numerals (see (§4.2.3.)).¹⁹⁰ This particular construction is used to indicate an approximate quantity rather than an exact number (and therefore, this is probably the motivation for the resulting syntactic agreement).

¹⁹⁰ T3 having produced the only exception to this rule in the entire corpus.

(115) a. (HL2.7 00:07) [HL3]

fiet-o moze bitj [...] **fia'dov** **lje...** s **prtdisjat** nazad,
that-NOM.SG.N maybe year-GEN.PL year from fifty-NOM ago

tʃɪ bolʃe

or more

‘That could have been around fifty years ago, maybe more.’

b. (B4.1 24:24)

njak-ɪ sovjet-ɪ nas sta-l-ɪ, da kroxi
some-NOM.PL Soviet-NOM.PL 1PL.GEN? appear-PST-PL and a bit

pobu-l-ɪ musjav **fio'dov** **dv-a** tʃɪ tr-ɪ
stay.PRF-PST-PL probably year-GEN.PL two-NOM.M or three-NOM

‘Some Soviets appeared, they stayed for a bit, around two or three years.’

c. (Tor1.24 05:00)

fiet-o uze minji bu-l-o **'rok-ɪv** moze
that-NOM.SG.N already 1SG.DAT be-PST-N.SG year-GEN.PL maybe

jak-ɪx **dvatsɪtj ʃɪstj,** **dvatsɪtj sɪm**
which-GEN.PL twenty-six.NOM twenty-seven.NOM

‘I was probably twenty six or twenty seven years old.’

It is interesting to note that the forms that are more rarely used in regular conditions are most often preferred in this context. For example, in the case of those varieties like B6's, who use the same inflection (and stem) for the NOMINATIVE SINGULAR and HIGHER NUMERALS, they use a different form with prenumerals (i.e. *rokɪv* and *fiodov*). This could point to the existence of an alternative GEN PL form for ‘year’ (a GREATER ADNUMERATIVE form). Hence, I find it more reasonable to classify those forms which are syncretic with the NOM SG as (dedicated) GREATER ADNUMERATIVE forms; and those

like *hodov* and *rokiv* as proper GEN PL forms (in the idiolects where there is such a distinction).

As a result, some of the morphosyntactic overabundance is reduced (fewer case suffixes available), but there seems to be no limit to the distribution of the suppletive stems.

6.2.4.2.4. Approximate quantity

Where there is uncertainty about the quantity, it can be expressed by the disjunctive conjunctions *ili*, *tʃi* and *abo* ‘or’. Since this is cross-linguistically common, some problems might be expected with the agreement. Nevertheless, in the utterances in the corpus, the agreement and the choice of stems do not suffer any consequence (at least when both numerals belong to the same class).

(116) a. (T3.2.2 01:35)

simnatsat ili vosimnatsat **lit.**
seventeen.NOM or eighteen.NOM year.GEN.PL

‘Seventeen or eighteen years.’

b. (T3.2.4 00:07)

odinatsat ili trinatsat **lit.**
eleven.NOM or thirteen.NOM year.GEN.PL

‘Eleven or thirteen [*sic*] years.’

6.2.4.2.5. With temporal adverbs

I have observed that, for certain speakers (such as B6), the suffix that they employ for the noun ‘year’ after a temporal adverb (such as ‘within’, ‘ago’) is special (i.e. different

from the forms they use in the other conditions considered here). This made me think that, perhaps, we could expect different results in terms of suppletive stems. Nevertheless, all three stems can appear in this context, with relative frequency.

(117) a. (T3.2.1 11:07)

tʃırız	pol	'fiɔd-a	bud-e	u	tebe	nov-i
in	half	year-GEN.SG	be.FUT-3SG	in	2SG.GEN	new-NOM.PL

‘In half a year’s time, you’ll have a new one.’

b. (T7.5 00:00)

[nɪ]skolko	lit	nazad
a_few	year.GEN.PL	ago

‘A few years ago.’

c. (HL2.32 02:07)

tʃerez	ɾik	propa-l-i	soovsem
in	year.NOM.SG	disappear-PST-PL	completely

‘They disappeared completely after a year.’

6.2.4.2.6. Perception of quantity

Besides the general syntactic rules applying to lower and higher numerals, I have not found any evidence for believing that the choice of one stem over another is conditioned by the subjective perception of the size of the amount (large or small), as I thought for it was for inflection.¹⁹¹ However, I have documented the three stems employed with the quantifier *mnũfiɔ* ‘many’.

¹⁹¹ In spite of the fact that, as I reported in the previous chapter (§5.3.), I had a suspicion for a long time that the use of the GREATER ADNUMERATIVE (instead of the GENITIVE PLURAL) was motivated by the perception of the amount (based on language consultants’ comments and perceptions).

(118) (Tor1.24 04:49)

[INT] a koljki v vas fiad-ov bu-l-o?
so how_many in 2PL.GEN year-GEN.PL be-PST-N.SG

[Tor1] to uze mnji bu-l-o **mnifio fio'd-ov**
so already 1SG.DAT be-PST-N.SG many year-GEN.PL

-[INTERVIEWER] So how old were you?

[Tor1] Oh, I was 'many years' old!

6.2.4.2.7. ANIMACY

Once again the corpus is too small to make a categorical claim. Nevertheless, it seems that there is no correlation between ANIMACY and the choice of the stem (e.g. 'This house is five years old' vs. 'The baby is five years old'). Most of the examples in the corpus with 'year' refer to people's age, although I have found a few instances of 'year' being employed to talk about NON-HUMANS.

(119) a. (Tor1.14 00:44)

sasudi-l-i na **simnatsat(j) rik** u tjurm-u,
sentence.PRF-PST-PL to seventeen year.GEN.PL/GRADNM in prison-LOC.SG

i t-oj K*** posidi-v **fest fiad**
and that-NOM.SG.M K. spend PRF-PST.M.SG six year.GEN.PL/GRADNM

'[He] was sentenced seventeen years, and this K*** spent six years in prison.'

b. (T1.4 01:25)

spirva mototsik [sic] jɪʃ, 'fɒda tri musit
first motorcycle.NOM.SG go.PST?.2SG? year.GEN.SG? three probably

‘In the beginning, I used to go on a motorcycle, for about three years,
probably’ [Obscure meaning].

6.2.4.3. Conclusions

To sum up, the putative conditions I have just examined reveal an even more complex distribution of the suppletive stems than we had seen in Table 29. So, if we look again at Table 29 we can see that there was not much homogeneity. Now, in Table 30, based on the same speakers as Table 29, after adding some conditions, not even a single column has a unique form, which points to OVERABUNDANCE (Thornton 2011, 2013) motivated by suppletion. However, this type of setting is to be expected from ‘real’ or naturalistic data. Furthermore, it must be recognised that part of the complexity is rather related to the unresolved morphological competition in this part of the paradigm. On the one hand, according to Thornton (2019) overabundance has its origin either in suppletion or in heteroclisys (i.e. two noun forms belonging to two different inflectional classes). The alternations or instances of overabundance between *fɒd* ~ *fɒdov* and *rɪk* ~ *rokɪv* (originally, inflectional class II) are definitely motivated by heteroclisys. At some point speakers may have followed the analogy of *lit-* (which being an inflectional class III noun, can have zero suffix in GEN PL / GRADNM) with higher numerals (and corresponding conditions) and got rid of the distinctive *-ov/iv* GENITIVE PLURAL suffix. Yet, even though Thornton (2019) only finds heteroclisys and suppletion as the logical source of overabundance, I dare to speculate that language contact in sociolinguistic contexts like

the one in Western Polesie can also be a possible source. On the other hand, if we leave the inflectional complexity aside for now, there are still some generalisations for predicting the stem that can be proposed, that decrease part of the suppletion-motivated overabundance:

- a) The stem *nik-* is dispreferred where it is not governed by a cardinal or collective numeral (except for 'one') or quantifier.
- b) The stem (and, most often, the form) used when the noun is in a pre-numeral position must be one that is available for higher numerals.
- c) It is possible to use exclusively *lit-* and *nik-* for certain cells, but there must always be a *hiod-* form available somewhere in the paradigm.

Table 30 Distribution of suppletive stems with additional conditions

	ORDINALS	NOM SG	LOWER ¹⁹² NUMERALS	HIGHER NUMERALS	PRENUMERAL	QUESTIONS, QUANTIFIERS
B6	fod	fod	'fodɪ	fod	fio'dɪv	fod
T3	rik	fod	'foda	lit	'foda ¹⁹³	
	fod					
HL4	fod	fod	'foda	lit	lit	
HL2			'foda	ljet	lit	
Tor1	fod	fod	'foda	rik	ro'kiɐ	'rokɪɐ
				'rokɪɐ		
				fod		fio'dov
				fio'dov		
Xab1			fio'da	ljet		
				ro'kiɐ		
TL6	fod		'foda	ljet	lit	fia'dov
			'rokɪ	ro'kiɐ		
HL3	fod	rik	fio'da	lit	fia'dov	
			'rokɪ	fio'dov		

¹⁹² Again LOWER and HIGHER NUMERALS here only covers those instances in which they appear in DIRECT CASES.

¹⁹³ This speaker produced a very peculiar prenumeral form nowhere else attested. We would expect some sort of GEN PL/ GRADNM to precede the numeral (even if it is a lower one), but he used this form (similar to an ADNUMERATIVE/ GEN SG) preceding the numeral 'three'.

6.3. The noun ‘person’

Up to now we have seen that the noun ‘year’ has particularly interesting suppletion patterns, which as I show in (§6.5.1.), are close to the canonical instance. Now I will explore another noun, which is also commonly suppletive in the Slavonic family, and even cross-linguistically.

Once I advanced on the transcription of the recorded corpus, I realised that the noun ‘person’ also displayed a peculiar behaviour in West Polesian, which differed from what I knew from its Belarusian, Russian or Polish cognates. I was aware of the alternation between the stems *tʃolovik-* and *ljud-*, in West Polesian. Yet, further on, I noticed that a third stem, *duʃ-*, was another juggling ball of the suppletive noun ‘person’, with cognates in other Southern Slavonic languages.

(120) a. (B20.17 00:25)

i jak **tʃolo'vik** ide notʃu obizatelno puʒaj-e
and when man.NOM.SG go-3SG at_night necessarily scare-3SG

‘And when a person/man (= someone) walks [over] he always scares them.’

b. (B20.19 01:51)

teper uʒe tak-ix **lud-ej** praktiʃeski, potʃt i ni-ma
now already these-GEN.PL people-GEN.PL virtually almost and NEG-HAVE

‘Nowadays there are hardly any people like this left’ [sorcerers].’

c. (T1.18 01:16)[...]

bo pʃat, sʃem **duʃ...** na sʃem **tʃolo'vik...**
as five seven person.GRADNM to seven person.GEN.PL/GRADNM

‘[...] because [there were] five, seven people... for seven people.’

Under this heading, firstly, (§6.3.1.) I present the noun ‘person’ and its behaviour with numeral phrases across the Slavonic family. Secondly, I present some ‘ideal paradigms’ of the three stems involved (§6.3.2.), which as happened with ‘year’ are very frequently combined and mixed by the speakers. In addition to this, (§6.3.2.1) the stems *ťolo'vik-* and *duf-* present complications, as there are homophonous forms existing with full paradigms. This will lead us again to propose some conditions in order to see whether the choice of one stem over another is restricted and/or motivated by them (§6.3.3.).

6.3.1. The noun ‘person’ across the Slavonic family

In order to make sure that I was not dealing with overlapping synonyms, I undertook another cross-Slavonic survey so as to see the behaviour of the noun ‘person’ (Table 31; Table 32; Table 33). In this survey, I included Macedonian (MKD) (as separate from Bulgarian) and Serbian (SRB) (as separate from Croatian) as I obtained significantly different results for these pairs. In addition, since the results of this survey are quite heterogeneous within each sub-family, I present them according to their genetic/areal affiliation. Where there have been many forms, I have stressed in bold the most common or dominant form. The areas in grey indicate the absence of data or results.¹⁹⁴

¹⁹⁴ Although the vast majority of results come from observations from the *ParaSol* corpus (Waldenfels & Meyer 2011), I had to use *Hrvatski Jeyzčni Portal* (<http://hjp.znanje.hr>) and *Rečnik na bālgarskija ezik* (<http://ibl.bas.bg/rbe/lang/bg/>) in order to confirm some of the results.

Table 31 Southern Slavonic

	NOM SG	LOWER NUMERALS		HIGHER NUMERALS	NOM PL
BG	čovek	čoveka		čoveka	čoveci
		xora ¹⁹⁵		xora	xora
		duši		duši	duši(te) ¹⁹⁶
HR	čovjek	čovjeka		ljudi	ljudi
SVO	človek	[2] človeka	[3] ljudje	ljudi	ljudje
MKD	čovek	čoveka ¹⁹⁷		duši	duši(te) ¹⁹⁸
		duši			
		luđe		luđe	luđe
SRB	čovek	čoveka		ljudi	ljudi

Table 32 Western Slavonic

	NOM SG	LOWER NUMERALS		HIGHER NUMERALS	NOM PL
CZ	člověk	lidé		lidí	lidé
POL	człowiek	ludzie / ludzi		ludzi	ludzie
SK	človek	ľudia		ľudí	ľudia
US	čłowjek	[2][n.d.]	[3-4] ludžo	ludži	ludžo

¹⁹⁵ Very marginal and it can only appear with an animate numeral.

¹⁹⁶ This form is very marginal and all the corpus results point out that, unless an article is used, the stem *duši* can only be used with quantifiers. Moreover the *Dictionary of the Bulgarian Academy (Rečnik na bǎlgarskia ezik, 2018 (online))* also notes that *duši* must be used with quantifiers.

¹⁹⁷ Only one hit in the entire corpus.

¹⁹⁸ As in Bulgarian, this form is marginal and can only appear with an article.

Table 33 Eastern Slavonic

	NOM SG	LOWER NUMERALS	HIGHER NUMERALS	NOM PL
BLM	čalavek	čalaveki	čalavek	ljudzi
			ljudzej	
CSR	čelovek	čeloveka	čelovek	ljudi
			ljudej	
ULM	ljudyna	ljudyny	čolovik	ljudy
			ljudej	

Due to time limitations, and in order to be coherent with the study done for ‘year’, I focus only on suppletion in contexts where there are numerals (i.e. where dedicated adnumerative forms are used),¹⁹⁹ and NOM SG – NOM PL, in order to see the contrast.

Whilst the noun ‘year’ in the Southern Slavonic family is very regular and unexciting; there is a very rich variation when it comes to the noun ‘person’. Other synonyms have been interfering with corpus results. The forms *osoba* (and the like) and *lica* (and the like) are present in many Slavonic languages alongside the more established forms (*človek*, *ljudi*, *duši* and the like). Hopefully, in most languages I have found enough evidence for stating that such forms are not additional stems, but just synonyms. For example, the following sentence would not be allowed in Slovak if the stems were suppletive (i.e. two suppletive stems in a disjunctive) “[...] *obyčajní ľudia či osoby nevedomé*” ‘[...] ordinary people or unknown persons’

¹⁹⁹ Regardless of the morphosyntactic strategy they follow; some may have a morphophonologically dedicated form, others may use the regular NOM PL form.

(Waldenfels and Meyer 2011: 77439)). After refining the results of the corpus, we can see that the suppletive stems under discussion (for West Polesian ‘person’) are present in other Slavonic languages, particularly in the Southern sub-family.

6.3.2. West Polesian suppletion patterns for ‘person’

Based on observations from the corpus, these are the underlying paradigms of the three stems for ‘person’. Note that for *lud-* many speakers produce it as [ljud-].²⁰⁰ Nonetheless, for the sake of consistency (and so as to merely focus on suppletion), I have kept the non-palatalised form, in the paradigm below (Tables 34 a, b, c). As I did for the noun ‘year’, I represent the cells containing those stems following a colour code: *ťolovik-*, red; *lud-*, blue; *duf-*, yellow.

The reader may have noticed that what I have included here (Tables 34 a, b, c) as the VOCATIVE (SINGULAR) form (*ťolo'vŕťe!*) can be either a derived form (with an appreciative), or a non-productive type of stem alternation. In favour of the second position, it must be said that the alternation between *-k* and *-ť* is almost certainly related to a phonological process from the Common Slavonic era known as the “First Palatalization” (Shevelov 1964). This is certainly an old remnant of that, especially taking into account that the VOCATIVE is a marginal CASE value and perhaps more susceptible to retain older forms (see the discussion on CASES in (§3.2.1.)). This also happens with some of the Belarusian and Russian remnants of the older VOCATIVE; e.g. (Russian) [NOM SG] *Bog* > [VOC SG] *Bože!* ‘oh, God!’.

²⁰⁰ And for most speakers there is free variation in the non-direct cases between the two realisations.

Tables 34 a, b, c Available stems for the noun ‘person’ in West Polesian

(a) the stem *tʃolovik-*

	SG	PL
NOM	tʃolo'vik	
VOC	tʃolo'vitʃe!	
ACC	tʃolo'vika	
GEN	tʃolo'vika	tʃolo'vik
DAT	tʃolo'viku	
INST	tʃolo'vikom	
LOC	[n.d]	
ADNM	tʃolo'viki	

(b) the stem *l(j)ud-*

	SG	PL
NOM		'ludi
VOC		'ludi!
ACC		lu'dej
GEN		lu'dej
DAT		'ludjam
INST		ludj'mi
LOC		'ludjax
ADNM		

(c) the stem *duf-*

	SG	PL
NOM		du'ʃɪ ²⁰¹
VOC		
ACC		
GEN		duʃ
DAT		
INST		
LOC		
ADNM	'duʃɪ	

6.3.2.1 Homophony and ambiguity

One of the biggest challenges for the analysis of the corpus has been the ambiguity caused by the polysemy (or homophony) of the stems *tʃolo'vik-* and *duf-*. The latter derives from *dufa* ‘soul’ (most likely motivated by metonymy). The form *duʃ(a)* can be also found unbound, but it denotes literally ‘soul (i.e. spirit)’.

Corbett (2007) and Vanhove (2017) describe a similar problem with a few Russian suppletive nouns such as *rebënok-detì* ‘child’. The noun *rebënok* is not available for the PLURAL; and so, for the rest of the forms, it uses the stem *det-*; i.e. [NOM PL] *detì*; but

²⁰¹ I have only recorded one instance of *duf-* being used (unbound) in the NOM PL, where it seems to mean ‘people’, instead of ‘souls’. See (§6.3.3.), for a further debate.

not **rebënki*. The noun *ditjo* ‘child’ exists on its own, but is rather archaic and restricted to the literature. As a result, the most ‘direct’ or semantically regular [NOM PL] of *rebënok* is *deti*.²⁰²

In short, the fact that *dufa* can stand on its own and has a meaning closely related to ‘person’, does not impair part of its paradigm from being used as a suppletive form of ‘person’ (rather than it being a mere synonym). Moreover, we have the cross-Slavonic survey (Table 31, Table 32 and Table 33) as evidence of this form being employed as a suppletive form of the ADNUMERATIVE forms, most remarkably in Bulgarian (Table 31). Consequently, from now on, whenever there is a possibility of confusion I will refer to *dufa*¹, as one of possible suppletive stems of ‘person’; and *dufa*², as a proper non-defective noun, meaning ‘soul’.

The forms of *tfolovik*- are far more complex. There is a continuous overlap of forms. Nevertheless, the noun *tfolovik*¹ ‘person’ is, at least, homophonous with *tfolovik*² ‘man, husband’, which I am also going to distinguish with supra-indexes for the sake of clarity.

(121) (B21.6 00:43)

tfolovik²	u	jijı	umer	u	sorok	pjat
husband.NOM.SG	in	3SG.GEN.F	die.PRF.PST.M.SG	in	forty	five.NOM
 hod	u	babı	ot	tif-u		
year.GRADNM	in	lady-GEN.SG	from	typhus-GEN.SG		

‘This lady’s husband (lit. man) died from typhus at the age of forty-five.’

²⁰² Some people may argue that this noun also has a NOMINATIVE PLURAL and VOCATIVE PLURAL form, based on truncation [NOM PL] *rebjata!*; [VOC PL] *rebjat!* Although these two forms may be originally related to the SINGULAR form *rebënok*, the correlation between the two is not semantically regular. The SINGULAR forms ought to be translated as ‘child’; whereas the PLURAL forms are an informal way of addressing teenagers or adults, and so they could be translated as ‘guys’.

The evidence for this distinction is that in some varieties these nouns inflect differently in the cell of the ADNUMERATIVE. When it means ‘man or husband’ it has a prototypical ADNUMERATIVE and GEN PL/GRADNM of inflectional class II, and syntactically it behaves as a PLURAL noun. As we have seen in example (83) in (§5.2.3.1.1.), or we can also see in (122).

(83) (T11.6 00:25)

tut	dva	tʃolo'viki,	dva	'xloptsɪ
here	two.NOM.M	man.ADNM	two.NOM.M	boy.ADNM

zabi-l-o	moljnija
kill.PRF-PST-N.SG	lightning(N).NOM.SG

‘The lightning killed two men, two boys here.’

(122) a. (T8.5 02:38)

muʒi'k-iv	vs-ix,	tʃolovi'k-iv²	ostavi-l-i	na	nitʃ,
male.GEN.PL	all-GEN.PL	man-GEN.PL	leave-PST-PL	in	night.ACC.SG?

notʃova-l-i	vonɪ	tam
overnight-PST-PL	3PL.NOM	there

‘All the males, the men were kept [there] for the night, they spent the night there.’

b. (T8.5 05:03)

nas	ni	pobi-l-i,	'ljud-aj,	nik-otio.	i	tʃolovi'k-iv²
1PL.ACC	NEG	beat.PRF-PST-PL	person-ACC.PL	nobody-ACC.SG	and	man-ACC.PL

tix,	muʒi'k-iv	vipusti-l-i	tʃerez	nitʃ
that-ACC.PL	male-ACC.PL	release.PRF-PST-PL	through	night.ACC.SG

‘And they did not kill any of our people; and those men, those males were released after the night.’

c. (Tor1.25 06:51)

tri	ʊzrozl-ix	tʃolo'viki	xodi-l-i
three.NOM	adult-GEN.PL	man.ADNM	go.IMPf-PST-PL

‘Three adult men were walking.’

I reiterate that such a distinction is not clear-cut and that for many varieties both lexemes inflect the same way.

6.3.3. Conditions for suppletion for the noun ‘person’

As I did with the noun ‘year’, I examine whether any conditions (semantic or syntactic) determine the choice of different stems. I tried looking at the same conditions as in ‘year’ (§6.2.4.), but not all of them are applicable to the noun ‘person’. Hence, I have omitted the ANIMATE/INANIMATE condition (as ‘person’ is inherently HUMAN ANIMATE); status numerals and temporal adverbs (since because of their semantic valency, they are only appropriate with nouns from the semantic field of TIME); and ordinal numerals (because they are irrelevant for this particular noun).²⁰³ I tried adding an extra condition: “the noun ‘person’ with collective numerals”; but, unfortunately, there are no instances of collective numerals + ‘person’ in the corpus.

6.3.3.1. Unbound or non-governed ‘person’

All three stems under discussion are possible, and as happened with the noun ‘year’, the suppletion patterns vary from one speaker to another. Yet, as far as I have been

²⁰³ They have the same effects on the noun ‘person’ as any other adjectives.

able to test, the noun ‘person’ uses at least two different stems in every speaker’s variety.²⁰⁴

(123) a. (Z1.6.1. 03:51)

dofanja-je nas **tsolo'vik**²⁷
get-3S.SG 1PL.ACC man.NOM.SG

‘The man gets us.’

b. (Z1.6.1. 05: 33)

ʃto 'ljudjam bud-e, te jej nam!
what people-DAT.PL be.FUT-3SG same 1PL.DAT

‘Whatever is to happen to the rest of the people, shall it also happen to us!’

When it comes to unbound forms, the stem *duf-* is certainly dispreferred, if not ungrammatical. This suggests that, most likely, it replicates the behaviour of its cognate in Bulgarian and Macedonian. It is possible to find the form *dufa*² unbound, but as I have said, it literally denotes ‘soul; spirit’. Not surprisingly it seems impossible to say something like ‘nasty people’ using *duf*⁻¹ (**biztolkovı dufı?*).²⁰⁵ In any case, I have only found two instances of unbound *dufa*¹ in the corpus (124), and both of them doubtful.

²⁰⁴ It is interesting to remark that in spite of the fact that the noun ‘year’ is more canonically suppletive and lexically more salient, it does not display suppletive paradigms for certain speakers (particularly in Bahdanaŭka). The noun ‘person’ being a lot more restricted in use and less canonically suppletive (as I show in (§6.5.)) displays a suppletive paradigm in the speech of every participant for whom there is enough data.

²⁰⁵ However, I have found a similar construction with the other two (see also (133) a., *infra*):

(TL6.3 01:59) priʃov starenjki **tsolo'vik**
‘An old man arrived.’

(Tor1.29 00:35) xoroʃtje tje 'ljudi
‘Those good people.’

(124) a. (B1.1 19:20?)

odno-jej **duʃɪ** ne pustɪ-ʋ
 one-GEN.SG.F person/soul?-GEN.SG NEG let-PST.SG

‘[He] wouldn’t let a single person in [lit. ‘not a soul’].’

b. (TL1.1 16:36)

njeskolko fio'd-ov sobira-l-i-sj, i **duʃɪ**
 some year-GEN.PL gather.IPFV-PST-PL-REFL and people.NOM.PL

[nu] sosjedje ʋ asnobnom, vsje prixodi-l-i
 well neighbour-NOM.PL in majority all-NOM.PL come.IPFV-PST-PL

‘[We] were gathering for several years, and people, that is to say, mostly neighbours, all used to come.’

In the first one (124) a., the stem *duʃ-* is not governed by a lower or higher numeral, the preceding constituent is a quasi-adjectival numeral (see Corbett 1978b). The second example (124) b. comes from another elderly speaker, whom I have recorded using the stem *duʃ-* with numerals as a suppletive form of ‘person’. However, it may be that the speaker really had in mind its primary meaning ‘soul’ when she produced this sentence.

6.3.3.2. Bound forms

Even more than with the noun ‘year’, most of the conditions which could have an effect on the distribution of the stems are those syntactic contexts in which ‘person’ is governed.

6.3.3.2.1. Bound to a cardinal numeral

When the noun ‘person’ is governed by a cardinal numeral all three stems can appear.

(125) a. (Tor1.25 07:50)

t-ix	Xavansk-ix	tože	admatsat	duŋ,
that-ACC.PL	Khavansk-ACC.PL	also	eleven	person.GRADNM

tʃilo'vjek	zavizli ²⁰⁶
person.GRADNM	bring.PRF-PST-PL

‘They also brought eleven people from Khavansk.’

b. (TL6.2 00:48)

dvatset	tʃolo'vik	v	komnat-i
twenty	person.GRADNM	in	room-LOC.SG

‘Twenty people in the room.’

c. (Tor1.47 01:47)

tam	milijon-i	lu'd-ej	zakopan-ix
there	million-ACC.PL	person-GEN.PL	buried-GEN.PL

‘There are millions of people buried there.’

6.3.3.2.2. With question words (Q) and quantifiers

When a question word (Q) or a quantifier is governing the NP, I have documented the stems *tʃolovik-* and *lud-*, and so far, I have not found the stem *duf-* used for this. Nevertheless, I do not see strong reasons why it could not be employed in these contexts as well.

²⁰⁶ Note that the speaker was aware of the suppletive stem *duf-*, but probably in an attempt to be understood by me, an outsider, she tried to simplify the paradigm for me.

(126) a. (Tor1.47 04:20)

z dirjevn-I vizva-l-I **mnifio** **lu'd-ej**
from village-GEN.SG call.PRF-PST.PL many person-GEN.PL

‘They called out many people from the village.’

b. (TL3.2 00:44)

njeskolko **tfolo'vik** povjerova-l-I
some person.GRADNM believe.PRF-PST-PL

‘Some people came to faith.’

c. (B19.3.0 00:26)

skaž-ete mnji kofo vi batʃi-l-i, **kiljka** **tfolo'vik**
tell.IMP-2PL 1SG.DAT Q.ACC 2PL.NOM see-PST-PL how_many person.GRADNM

vi batʃi-l-i, xto bud-e, xto **bilʃ** ufiljed-it
2PL.NOM see-PST-PL REL.NOM be.FUT-3SG REL.NOM more see.PRF-3SG

lju'd-ej tomu ja spodnits-u kup-lju
person-GEN-PL REL-DAT.SG 1SG.NOM skirt-ACC.SG buy.PRF-1SG

‘Tell me how many people (men) you saw. Whoever sees more people, I will buy her a skirt.’

6.3.3.2.3. With the adjective ‘all’

With the adjective *vsi* ‘all’ only the stem *lud-* has been attested, and it seems unlikely that *tfolovik-* or *duf-* would be allowed in such a context.

(127) (HL2.16 02:21) [HL4] [...]

vs-ix lu'd-ej v adn-u kup-u [sic]
all-ACC.PL person-ACC.PL in one-ACC.SG pile-ACC.SG

‘[...] all the people in one pile.’

6.3.3.2.4. Before the numeral

As I have already explained for the noun ‘year’ (§6.2.4.2.3.) and in my introduction to numeral phrases (§4.2.3.), when a cardinal numeral is used after the noun it determines, it indicates that the quantity is approximate (rather than exact). In these instances the NP stands in GEN PL (or GR.ADNUMERATIVE), in spite of the fact that the numeral may be a LOWER one. When it comes to suppletion, so far, I have only documented *tʃolovik-* and *duf-*, but not *lud-*. This restriction is also shared by Russian, so there seems to be evidence to believe that the stem *lud-* is ungrammatical in this type of construction.

(128) a. (B20.8 00:11)

zbira-l-a-sa kompan-ja; tʃolo'vik deset, dvanatsat
gather-PST-F.SG-REFL company-NOM.SG person.GRADNM ten twenty

pri... primjerno odn-ofio fiod-u.
for_example one-GEN.SG.M year-IIGEN.SG

‘A group used to gather, about ten people or twelve people of more or less the same age (lit. year).’

b. (T1.3 03:18)

tʃtɪrɪ	stol-a,	tʃolo'vɪk	ʃestdisjat,	vosimdisjat
four.NOM	table-GEN.SG	person.GRADNM	sixty	eighty

i	tʃtɪrɪ	dnj-a.
and	four.NOM	day-GEN.SG

‘Four tables, around seventy or eighty people and four days.’

c. (TL1.1 16:45)

duʃ	djesjet	navjerno, [...] ja	tak	prɪdpolofia-ju
person.GRADNM	ten	probably	1SG.NOM	so guess-1SG

duʃ	djesjet	pokajɪ-l-o-sja
person.GRADNM	ten	repent-PST-N.SG-REFL

‘About ten people, most likely [...] I guess that around ten people repented.’

6.3.3.2.5. Approximate quantity

When an approximate quantity (rather than an accurate cardinal form) is used, the stem *lud-* is not permitted.

(129) a. (Z4.1.1 09:04)

xo'dɪ-l-ɪ	po tri,	po pjat tʃolovɪk	i	ʃ-l-ɪ	u lis
go.IPFV-PST-PL	in three	in five person.GRADNM	and	go-PST-PL	to wood.ACC.SG

‘[They] used to go to the forest [in groups of] three-five people.’

b. (T1.1 24:03)

tam	bu-u	tʃolovɪk	moʒe	tɪsɪtʃa
there	be-PST.M.SG	person.GRADNM	maybe	thousand

‘There were about three hundred people.’

c. (T1.18 01:16)

pjat, sjem duʃ, na sjem tʃolo'vik
five seven person.GRADNM for seven person.GRADNM

‘Five or seven people, [enough] for seven people.’

6.3.3.2.6. Perception of quantity

I have found instances of the stems *tʃolovik-* and *lud-*, with larger and smaller numbers, but not *duf-*, probably because there are not many instances of the stem *duf-* in the corpus. However, there is no evidence that the choice of the stems is related to any pragmatic/semantic factors; i.e. whether the speaker perceives the number of people as small (and thus, easily individuated) or large. In (130) I show some examples of all the stems being used with large amounts.

(130) a. (T11.2 00:31)

bu-l-o po sto pʁtdɨsʃat, po dvjestɨ tʃolo'vik
be-PST-N.SG about hundred_and_fifty about two_hundred person.GRADNM

na svadv-e ʃulja-l-o kolɨsj...
in wedding-LOC.SG party-PST-N.SG formerly

‘About 150-200 people used to party in [each] wedding.’

b. (Tor1.6 00:50)

vin tɨsatʃɨ lu'd-ej [...] pap-a spas
3SG.NOM.M thousand person-GEN.PL father-NOM.SG save.PRF.PST.M.SG

‘[It can be said] that [my] father saved a thousand people.’

c. (TL1.1 17:43)

djevjenosto dvje du'ʃɨ bu-l-o v dom-i
ninety-two.NOM.F person.ADNM be-PST-N.SG in house-LOC.SG

‘Ninety-two people were at home.’

6.3.3.3. Sociolinguistic remarks

I have only been able to document the stem *dufa* (when it is used as a suppletive form of the noun ‘person’) in the speech of the older generation (born before 1950): B1, T1, TL1, Z4 and Z10, who also happen to be all female, except for T1. Nevertheless, my intuition is that the fact that only women appear using that form in my corpus is either accidental or motivated by an unbalanced gender sample,²⁰⁷ especially when it comes to the oldest speakers. Hence, the alternation between *duf-* and the other two stems seems more strongly conditioned by age, than for ‘year’ (for which sociolinguistics does not play any role).

I have not been able to find any instances of *lud-* in TL1’s recordings. This could make us suspect that she does not use the stem at all, although it seems very unlikely. I used to have a similar situation with Z4, who used *dufa* (as a suppletive root) in one of the sessions, but I have eventually been able to document a token of *lud-*, in an OBLIQUE CASE (131). Compare (17) from (§4.3.1.) to (131) and both produced by Z4.

(17) (Z4.1.2 00:49)

z nas jest duʃ: bu-l-o dɪt-ej ʃestero [...]
from 1SG.GEN six person.GEN.PL? be-PST-N.SG child-GEN.PL six(COLL)

‘Altogether, we were six people: there were six children [together] [...].’

(131) (Z4.1.1 03:41)[...]

iti po 'ludj-ax, moʒe tam komu ʃtʃjo pomofi-tɪ
go-INF for person-LOC.PL maybe there Q.DAT Q.ACC help-INF

‘[...] going to different people, maybe someone could help.’

²⁰⁷ More on gender imbalance of the sample in (§2.3.4.).

Thus, at least we can affirm that, although a minority, the three suppletive stems are present in some idiolects. And by looking at their behaviour in other Slavonic languages, we have more evidence to state that there are not overlapping synonyms, admitting that some stems are more frequent than others.

6.3.3.4. Summary

After applying some of the conditions, this is the resulting picture (Table 35). In comparison to the noun ‘year’ we can observe more consistency among speakers, which points to a more defined morphosemantic pattern (and thus less canonicity).

There are a few conclusions that can be drawn after having applied the putative conditions and comparing the results of different speakers:

- The noun ‘person’ has a suppletive paradigm in every speaker’s idiolect, which is usually composed of two stems.
- The stem *tfolovik-* appears in the NOM SG of every idiolect, but cannot appear in the NOM PL.
- The stem *lud-* is neither allowed with approximate quantities nor in a prenumeral position.

Table 35 Distribution of the suppletive stems based on utterances from speakers

	NOM SG	LOWER NUMERALS	HIGHER NUMERALS	NOM PL	PRENUMERAL POSSITION	+ QUANTIFIER OR Q WORD
Z10	tʃolovik ²		duʃ			
Z4			tʃolovik			
			duʃ			
TL1	tʃilovjek	'duʃɪ		duʃɪ	duʃ	
T1	tʃolovik	tʃolovikɪ	tʃolovik	ludɪ	tʃolovik	ludej
		tʃolovika	duʃ			
B20	tʃolovik		tʃolovik	ljudɪ		
Tor1	tʃolovik	tʃolo'vikɪ	tʃolovik	ludɪ	tʃolovik	ludej
			ludej			
T11		tʃolovikɪ	tʃolovik			ludej
TL3		tʃolovika	tʃjelovjek	ljudɪ		tʃolo'vik
			tʃolovik			
TL4	tʃolovik		tʃolovik	ludɪ		
				ljudɪ		
HL2		tʃolovika		ludɪ		

6.4. Challenges to the traditional assumptions about suppletion

Having presented the data from the corpus and addressed the conditions which may be affecting the choice of one stem over another, I have proved that in most cases, particularly for the noun ‘year’ there is simply overabundance. Thus, I am going to challenge some of the traditional assumptions about suppletion (which were already partly criticised in (Chumakina et al. 2004, Corbett 2007)).

Firstly, even though probably no modern-day committed linguist would state this, in the past some people (e.g. Osthoff 1899) believed that suppletion is a side effect of defectivity. That is to say, they believed that the main motivation for suppletion is to fill the holes of defective paradigms. In other words, they believed that the origin of suppletion was lexically motivated. Taking the example from the beginning of this chapter, the form *better* would be a ‘patch’ of the comparative form of *good* (which otherwise would be impossible; **gooder*). One could object that most of the PLURAL sub-paradigm of *tʃolovik*- ‘person’ is a counter-example to my argument. Nevertheless, a deeper analysis shows that the additional lexical stems and forms of ‘person’, and particularly ‘year’, do not fill any gap in West Polesian. They are rather an intrusion into the system, in part (probably) motivated by the contact with neighbouring varieties; and so there is no lexical motivation at all.

Secondly, another commonly held conception of suppletion has been depicting it as an ‘armistice’ to set peace between competing lexical stems or paradigms. For example, at some point in history, there were two (etymologically unrelated) verbs which had ended up meaning ‘to go’. Having arrived at this state, some theories, such

as “Gaussian Competition” (Aronoff 2016), roughly say that the competing stems have two options: they either specialise/resemanticise (e.g. one ends up denoting ‘to go by a means of transport’, and the other one ‘to go on foot’); or ‘split the cake’ (e.g. one stem takes the PRESENT TENSE paradigm; and the other the PAST).²⁰⁸ Suppletion would be the resulting state of that split. However, the cases of West Polesian ‘year’ and the GEN PL/GRADNM cell of ‘person’ show that different stems can ‘coexist happily’ without having to come up with an ‘armistice’.

Thirdly, it has also been assumed that suppletion aligns with a major semantic category (e.g. NUMBER: SINGULAR VS PLURAL). According to Bybee (1985),

“[...] suppletive paradigms are divided along the category lines that involve the greatest change in meaning [...] inflectional splits are most likely to coincide with distinctions in the more highly relevant categories, those that make the largest semantic change” (Bybee 1985: 92).

For this reason, for Bybee (1985), suppletion is far more likely to happen in verbal paradigms than in nominal paradigms. Nonetheless, the data for ‘year’ in West Polesian show how suppletion does not easily align with NUMBER or CASE. And the same can be said about the stem *dufa-* for ‘person’. Thus, this is not straightforward semantics (or morphosemantics) because the arbitrariness of the distribution of the stems (i.e. patterns) is more related to the peculiar morphosyntactic phenomena around quantification than to anything else.

²⁰⁸ Special thanks to Helen Sims-Williams (p.c.) for suggesting this point.

6.5. The nouns ‘year’ and ‘person’ in the light of Canonical Typology

In the previous section (§6.4.), we have seen that the traditional definitions or assumptions about suppletion needed revision in order to adjust to the reality found in the nouns ‘year’ and ‘person’ in West Polesian. For this reason, I will use the framework of Canonical Typology, more specifically, the typology or criteria to define suppletion provided by Corbett (2007).

Canonical Typology (CT) seeks to define a:

“CANON, a reference point from which to compare linguistic objects and descriptions. [...] It ensures linguists are talking about the same thing when comparing languages or structures, and avoid[s] problems associated with terminology [...]” (Bond 2018: 410-411).

Although the framework of Canonical Typology (CT) can also be applied to study the morphology-phonology interface (Hyman 2012), most of the work published on CT is related to inflectional morphology. Within the domain of inflection, the biggest (or at least the most studied) topics in CT are the non-canonical phenomena; i.e. deviations from regular, complete and ‘well-behaved’ inflectional paradigms. To mention some of the possible deviations studied: heteroclisis (Stump 2006); syncretism (Baerman et al. 2005); overabundance (Thornton 2011)); and suppletion (Corbett 2007).

Since in the West Polesian nouns ‘person’, and particularly ‘year’ different non-canonical phenomena come together I feel that CT provides a good theoretical and analytical framework. The most notorious non-canonical phenomenon in the available paradigms is suppletion, which is the main topic under discussion here. Nevertheless, there are also

instances of heteroclisis (with the forms *fi*od-; *ri*k-; and *tfo*lovik- , behaving like inflectional class I or III) in the GENITIVE PLURAL cell(s)); and overabundance.

Therefore, in the following section I analyse the nouns ‘year’ (§6.5.1.) and ‘person’ (§6.5.2.) in West Polesian, following the fourteen criteria Corbett (2007) sets in his article; and I finish with a summary of the results (§6.5.3.). His typology seeks to explore the canonicity of a non-canonical parameter. Corbett (2007) understands “more canonically non-canonical” as a ‘perfect irregularity’, in this case, the most extreme case of suppletion that can be theoretically proposed, even if no word in any language in the world can ever meet all the criteria.²⁰⁹

6.5.1. The noun ‘year’

CRITERION 1: Fused exponence

According to Corbett (2007: 15), “[...] if the suppletive form combines stems and inflection, this is ‘more canonically suppletive’ [...] than if the form is the stem to which appropriate inflections are added”.

In the case of ‘year’, it is not a straightforward question because suppletion behaves differently in different cells. On the one hand, the forms *fi*od- and *ri*k- can take GEN PL/GRADNM forms which are impossible to differentiate from NOM SG. As a result, the segmentation of their forms can be said to be opaque (synchronically). Nevertheless, the form(s) can also be analysed as having a zero/unmarked suffix

²⁰⁹ Or Bond (2018: 411) puts it this way: “[s]ince the canon is only a construct, a real life exemplar of it may not exist”.

(-∅), which some inflectional class III and most commonly class I nouns take for the GEN PL. Both *hōd-* and *rīk-* belong to inflectional class II (at least in the SINGULAR sub-paradigm). Thus this would comprise an interesting case of heteroclisis, but not fused exponence.

On the other hand, in the rest of the cases, *hōd-* and *rīk-* inflect like regular inflectional class II nouns, and so it is possible to distinguish the stems from the suffixes. Furthermore, the cell for higher numerals in many speakers' idiolects can be filled with *hōdov* and *rokuv*, which display the type of GEN PL we would expect from inflectional class II nouns.

The stem *lit-* also presents some challenges for its classification. Synchronically, the noun *lito* 'summer' exists independently in West Polesian. But in this particularly morphosyntactic context, it replaces the GEN PL of a completely different noun (synchronically). Except for one instance in the whole corpus, where I have identified the use of *lit-* in LOC PL, this form only appears in GEN PL / GRADNM. Thus, some could treat the form *lit-* as inflectionally unsegmentable (i.e. fusional). Yet, as I have just pointed out for *hōd-* and *rīk-*, the noun *lito* being clearly inflectional class III, and knowing that class III nouns can take either *-ov/iv* or *-∅* in these cells (e.g. as in *jaj'ts-iv* or *jajits* 'egg') the form is better analysed as displaying exponence.

Therefore, a careful analysis of the noun 'year' according to this parameter shows that it does not meet the canon.

CRITERION 2: Phonological distinctiveness

According to Corbett (2007: 15) in the canonical instance “the formal correlation is maximally irregular”. Although Corbett (2007) shows some concerns about the measurability of this criterion, the idea is that the more phonologically distinct the suppletive form(s) are the more “canonically suppletive” they are. Moreover, Bybee’s (1985) definition of suppletion only concerned those instances where the stems have different etymologies; which for Corbett (2007: 15-16, 24) is not a requirement. It is clear that in West Polesian there is no phonological or morphological rule which could derive *lit-* from *hōd-* or *rik-* (and viceversa). Moreover, it is solidly historically attested that the forms *hōd-*, *rik-* and *lit-* have different etymologies (Boldyrjev et al. 2006, Brückner 1927, Chumakina et al. 2004), so there is sufficient evidence to state that there is no formal correlation between them. Thus, there are no doubts for affirming that all the suppletive stems of ‘year’ fulfil this criterion of canonicity.

CRITERION 3: No overt realisation

According to Corbett (2007: 16-17), if one of the stems in the paradigm is not overtly realised (phonologically), this form is more canonically suppletive than those that are phonologically realised. That is to say, with the noun ‘year’ the CASE/NUMBER suffix would have to stand on its own, whereas other forms of the paradigm would consist of the suppletive stem + CASE/NUMBER suffix. Yet, there are no such forms like ‘*pjet* **ø-ov*’ to denote ‘five years’. Consequently, the noun ‘year’ fails to meet this canonicity criterion.

CRITERION 4: More than one variant

According to Corbett (2007: 17) “the higher the number of different ways in which the common lexical material is reflected, the greater the formal irregularity and the more canonically suppletive the lexeme”. As I have shown in Table 29, there are three different stems for ‘year’ in many speakers’ idiolects. That makes the West Polesian noun ‘year’ the most canonical of all Slavonic varieties, which either display no suppletion, or at most a two-way stem alternation (Czech, Polish and Russian).

CRITERION 5: Morphological distribution

According to Corbett (2007), the most “canonical instance of suppletion” is one in which the distribution of the suppletive stems across paradigms is motivated by a morphological pattern (or “morphomic”, in Aronoff’s (1994) terminology), rather than a morphosyntactic (perhaps, also morphosemantic)²¹⁰. For example, if the distribution of suppletion aligns with a major semantic category such as **NUMBER**, suppletion is contributing to the distinction between **SINGULAR** and **PLURAL**. As a result, just by hearing the stem of the **PLURAL** (without even hearing any other suffixes or prefixes) one can unequivocally know that the interlocutor is referring to a **NON-SINGULAR** entity. This means that the stem alternation is adding an extra semantic nuance; and thus the semantic regularity between the two forms is not perfect.

Corbett (2007: 19) proposes French as an example of morphological rather than morphosyntactic distribution of suppletion. In most French suppletive paradigms 1PL

²¹⁰ See a more detailed explanation in Kibort (2008).

and 2PL contrast with the rest of the cells in the PRESENT TENSE.²¹¹ This pattern, although recurring, is hard to define by a single morphosyntactic feature, a complex explanation or combination is needed.

On the one hand, if we disregard speakers' preferences or frequency of tokens, I have shown enough evidence of the stems *rik-* and *hod-* being available for any cell in some idiolects. Thus, they would not be following any predictable pattern. However, if the two lexemes' inventory of forms is 'complete', then they cannot be morphomic.²¹²

On the other hand, the stem *lit-* appears only in the GEN PL / GRADNM cell for most speakers. The only noun which remotely resembles it is the noun *tfolovik* 'person'. I have already introduced this noun, and I will give more details on this soon (§6.5.2.). Very briefly, this noun has special adnumeratives [ADNM] *dufi* and [GEN PL / GRADNM] *duf* (or *tfolovik*) for a handful of speakers, although all the remaining distribution of suppletive stems in the paradigms is completely different for both nouns. Just one noun is not enough material to comprise a pattern. Hence, there is little or no evidence to propose any distribution pattern for suppletive stems, but if there were one, it could definitely not be a morphosyntactic one, and thus, for this criterion, the noun 'year' is consistent with the canon.

²¹¹ For example, compare the PRESENT TENSE of the verb *aller* 'to go' in the PRESENT TENSE: [1SG] *je vais*, [2SG] *tu vas*, [3SG] *il/elle va*, [3PL] *ils/elles vont*; BUT [1PL] *nous allons*, [2PL] *vous allez*; with the PRESENT of the verb *avoir* 'to have': [1SG] *j'ai*, [2SG] *tu as*, [3SG] *il/elle a*, [3PL] *ils/elles ont*; BUT [1PL] *nous avons*, [2PL] *vous avez*.

²¹² Many thanks to Borja Herce (p.c.) for this explanation.

CRITERION 6: Alternating

According to Corbett (2007: 23-24), the most canonical instance of suppletion would involve the possibility of having multiple suppletive stems within the same cell (or cells). Such a criterion is better summarised by the concept of OVERABUNDANCE which was later coined by Thornton (2011) (and now used by other authors as well (e.g. Cappellaro 2013)). As has been proven in (§6.2.4.), suppletion patterns for the West Polesian noun ‘year’ fail to be regular from one speaker to another, and even within many speakers’ idiolects multiple stems are possible for a single cell. For example, in Table 29 the speaker Tor1 presents two different stems with higher numerals, whereas HL3 uses two different stems with lower numerals, and two others (in a different combination) for higher numerals. A more detailed analysis of the canonicity of this instance of overabundance could be made, but I try to limit the study to suppletion. In any case, data from West Polesian strongly suggest that according to this criterion, the West Polesian noun ‘year’ meets the canon.

CRITERION 7: Less relevant features involved

Bybee (1985: 76, 93) proposed that we are more likely to find suppletion in verbal paradigms than in noun paradigms. In the same vein, she suggested that we are more likely to find a lexical split (i.e. suppletion) along the lines of VALENCY, TENSE or NUMBER than for PERSON because of their semantic saliency. Moreover, she argues that we are very unlikely to find suppletion for CASE “since case does not affect the meaning of the noun stem, but only signals its relation to other constituents in a particular sentence” (Bybee 1985: 93). Corbett concludes that

“the more relevant the feature, the lower the semantic regularity, since the interaction with the semantics of the root is greatest [...] Therefore the less relevant the feature involved, the more canonical the suppletion” (Corbett 2007: 24).

The case of the West Polesian noun ‘year’ is especially non-canonical in this respect. On the one hand, in those varieties where there is a special form for the SECOND GEN PL (or GRADNM) (as the paradigm of Russian or Polish), suppletion is codifying a split involving more than one feature (see *CRITERION 9, infra*), NUMBER being a slightly more regular or common feature and CASE being one of the rarest (thus more canonical). On the other hand, in those varieties where more than one form is used outside the SECOND GEN PL (or GRADNM) cell (for higher numerals), suppletion does not seem to be strictly codifying any features. Therefore, this makes the West Polesian ‘year’ a good instance of canonical suppletion.

CRITERION 8: Contextual features

Corbett (2007: 24-25) (based on Booij’s work (1994, 1996), see references there) distinguishes between inherent inflection (i.e. happening at the level of the word, such as NUMBER), and contextual inflection (i.e. the one “dictated by syntax [...]” such as “structural case markers on nouns”. According to Corbett (2007: 25) “suppletion according to less relevant and/or contextual features creates greater opacity and is, therefore, a more canonical [...] instance of suppletion”.

The noun ‘year’ would fall under the criterion of being contextually determined in all the Slavonic languages where it is suppletive. In the case of those varieties where the

forms *fi*od- and *ri*k- are used across the cells of the paradigm, it would not be very clear whether that would be a syntactic condition or an inherent feature (as it would not be indicating any SG-PL alternation). However, in the documented varieties there is a general tendency to use *fi*od- with ‘one’ or in unbound forms, whereas *ri*k- tends to start appearing after lower numerals.

Except for the single instance of *lit*- in LOCATIVE PLURAL (111) a., *lit*- could be the only stem with a more inherent feature. But there are no more instances of this stem being used in the PLURAL (or even SINGULAR) outside the GEN PL / GRADNM context, so as to contrast it. Yet, even though there could be some NUMBER alignment, its CASE distribution still shows a contextually determined suppletion. In short, the noun ‘year’ fulfils this canonicity criterion.

CRITERION 9: More than one feature

Where the suppletive stem applies to a combination of features (e.g. 1SG.PRS) rather than a single one (e.g. NUMBER) the resulting type of suppletion is semantically more opaque and in this way more canonical, according to Corbett (2007: 25).

When it comes to the West Polesian ‘year’, in those idiolects in which *lit*- (very often) and/or *ri*k- (less frequently) are exclusively used as SECOND GEN PL (or GRADNM), suppletion is codifying more than one feature (CASE and NUMBER), which makes it more canonical. Some may argue that, in those idiolects where *ri*k- and *fi*od- can be employed in any cell of the paradigm, there are no clear features involved, making them less canonical as a result. Nonetheless, although *ri*k- and *fi*od- combine quite freely, the uses of *ri*k- are more restricted for most speakers. Moreover, there are no

clear instances of suppletion exclusively applying to a **NUMBER** or a **CASE** value (for example *lit-* lacks an **ADNUM** (SG)/**GEN** SG counterpart). Thus, the noun ‘year’ is canonical in this respect.

CRITERION 10: No overlapping

Where one or more of the segments of the suppletive paradigm are being imported from another paradigm (e.g. from a different verb or noun) an **OVERLAP** takes place. Corbett (2007: 26) considers that such overlap makes “the semantic regularity within a single paradigm less clear”, and in this respect, less canonical. But that overlap can sometimes be tracked **synchronically**, causing a **DIRECTIONAL OVERLAP**.²¹³ That is to say, the paradigm where the suppletive form comes from can still be recognised, and therefore it can be deduced which paradigm is taking over which (and thus, semantic regularity is affected). According to Corbett (2007: 26-27), directional overlaps are less canonical than non-directional overlaps.

On the one hand, if we did not have any information about the etymology, it would seem that different segments of the paradigm of ‘year’ overlap with these of ‘summer’ in Contemporary West Polesian. Yet, I have already explained in ((§6.2.2.) based on Chumakina et al. (2004: 290)), the form *hod-* is an intrusion in the paradigm of *let-*, which was the main form in CES, and not vice versa. So, in this respect, there is an overlap (and thus, it does not meet this canonicity criterion), but this overlap, not being clearly directional, is closer to the canonical instance.²¹⁴

²¹³ I shall emphasise that for Corbett (2007: 26-27) “overlapping” is only to be determined **synchronically**.

²¹⁴ The same applies to Russian, Polish and Czech.

On the other hand, the overlap of the stems *fi*od- and *ri*k- is undeniable. However, it is not clear which one has taken over which. Moreover, even though this is a synchronic criterion, both forms have their tradition in other Slavonic languages. In the case of the West Polesian varieties in which *ri*k- is used, it may be that *ri*k- was the most widespread form and then the *fi*od-/lit- system intruded (perhaps due to the contact with Russian or other varieties where *fi*od- is used).²¹⁵

One could still argue that, cross-linguistically, overlaps tend to follow a morphosemantic pattern (in this case, most likely NUMBER).²¹⁶ The form *ri*k- seems rarely available for the SINGULAR (excluding the ADNUMERATIVE which has a very complex morphosyntactic status (Chapter 5)). However, cases like the speaker TL3, who has been documented using *ri*k- with an ordinal numeral, or HL3, who uses the form *ri*k- with ‘one’, make such a claim problematic. Furthermore, since the existing individualised corpus is very short for most speakers (in most cases up to 30 minutes of recorded texts), there is the possibility that the form *ri*k- could appear in more contexts (including SINGULAR). Therefore, reconsidering the critique to the necessity of semantic alignment for suppletion in (§6.4.), the distribution of *fi*od- and *ri*k- does not follow any morphosemantic pattern (see CRITERION 5, *supra*), and thus, it has an apparently free distribution.

²¹⁵ The same argument could be applied in reverse; originally *fi*od- or lit- were used in West Polesian, and the form *ri*k- (PL *rok*) appeared under the influence of Polish in the region. However, in the corpus *ri*k- tends to appear with numerals (especially with higher numerals) and quantifiers, which Polish avoids (i.e. *Ile masz lat?*, but **Ile masz roków?* ‘How old are you?’; *ma dwa lata*; but **ma dwa roka* ‘s/he is two years old’).

²¹⁶ Corbett (2007: 26) already showed a counter-example to that tendency with the paradigm of Latin [NOM SG] *nemo* [GEN SG] *nullius* ‘nobody’.

To sum up, the noun ‘year’ (in whichever combination of stems) shows an overlap in its paradigm(s), and thus, it is less canonical in this respect. Nevertheless, given that the directionality of this overlap is certainly unclear synchronically, it is partly canonical.

CRITERION 11: Absence of remainders

In Corbett’s (2007: 26) understanding, “remainders” are those instances in which the forms supposed to be replaced by the suppletive stem are still in use.²¹⁷ He treats such instances as less canonical than the ones without remainders as the remainders have a semantic correlation with the “main form”.

When it comes to West Polesian, those varieties that have the forms *fi*od- and *ri*k- spread all over the paradigm are the least canonical of all, in this respect, because it is not clear which form is the proper suppletive form and which one is the remainder. And looking at *li*t- as the ‘real’ suppletive form (from a synchronic perspective), it still has remainders in many idiolects. Yet, once again, more data is needed in order to state that all those idiolects for which no remainders have been documented in the corpus do indeed lack remainders. Hence, those varieties which have remainders alongside the form *li*t- (or less rarely *ri*k-) are less canonical than those which do not. In general, it can be said that the noun ‘year’ fails to meet this requirement to be canonical.

²¹⁷ Although in cases like ‘year’ in Russian, in which the suppletion is the result of an intrusion (i.e. as I pointed out in (§6.2.2.) Chumakina et al. (2004: 290) affirm that the form *let* used to exist, before *god* gradually took over the paradigm), the remainders would not be a vestige of an older form, but the result of an unfulfilled invasion. Therefore, *godov* would not be the root that had to be replaced *stricto sensu*, but the one which hardly managed to get into the system. However, for the sake of practicality, I am also classifying them under the umbrella of ‘remainders’.

CRITERION 12: Uniqueness (non-productive derivationally)

According to Corbett (2007: 27) “[i]f the formal correlation is repeated, the phonological relation is no longer absolutely irregular”. That is to say, if there were derived forms of *ɦod-*, *rik-* or *lit-* in West Polesian replicating the suppletive pattern, it would make them a less canonical type of suppletion. Nevertheless, the only derived forms of *ɦod-* I can think of are *ɦodikov*, GEN PL of *ɦodik*, which also means ‘year’ in child-speech (i.e. (B6.ov?) *kɫjke toɫ ɦodikov?* ‘how old are you?’) and perhaps the verb *ɦodovati/hudovati* (i.e. to keep or take care of an animal, most commonly a pig; usually for an entire ‘year’, which may be the etymology). With respect to *ɦodikov*, it is a loan from Russian (used in the same context and with the same restrictions). In any case, it is impossible to say **litkov* or **litovati*.²¹⁸ The suppletive stems for ‘year’ in West Polesian are unique or limited from the point of view of morphological (derivational) productivity, and consequently more canonical in this respect.

CRITERION 13: No outside conditions

According to Corbett (2007: 27-29) instances of suppletion that only appear under certain conditions (e.g. “phrasal suppletion”) are less canonical than those in which suppletion is not affected by external conditions.²¹⁹

This criterion is tricky with the noun ‘year’. If we are to analyse the adnumerative forms as CASE or even NUMBER values, then we can affirm that there are no outside

²¹⁸ And my intuition says that you cannot equally say something like **rokɫkv*; however, I have never tried to elicit that form from any speaker.

²¹⁹ The author makes a clear distinction between proper *outside conditions* and *shape conditions*, the last being the result of a morphophonological process.

conditions (other than the ones normally applied to *CASES*). In other words, the suppletive distribution is already in the paradigm, independent of the numerals governing the phrase. However, the *ADNUMERATIVE* cell(s), particularly the one that I call (LOWER) *ADNUMERATIVE*, can only be used if there is a specific type of numeral governing the phrase. In any case, in the previous chapter (§5.6.) I had concluded that the best analysis of West Polesian adnumerative forms, or at least the (LOWER) *ADNUMERATIVE*, is that it is a hybrid feature value “in the penumbra of feature values” (rephrasing Corbett (2011)), which is integrated in the paradigm.

Furthermore, I have provided enough evidence in (§6.2.4.) to demonstrate that syntactic or semantic conditions have little or no effect on the choice of a particular stem. Thus, according to this parameter, the noun ‘year’ is canonical.

CRITERION 14: Absence of syntactic effects

In spite of the features certain cells have (e.g. *PLURAL*) that may affect agreement, if suppletion affects the syntax in another non-predicted way (e.g. different stems governing different *CASES*) there will be less evidence of semantic regularity, according to Corbett (2007: 29). Hence, instances of suppletion that trigger different syntactic effects are less canonical, as they are more likely to belong to different paradigms (i.e. synonyms). None of the suppletive forms in West Polesian present outgoing or external syntactic effects. Thus, in this respect, this suppletion meets the criterion to be canonical.

Summary of the criteria

The results from these canonicity criteria can be summarised (using a binary approach) in Table 36:

Table 36 Summary of the canonicity criteria for the noun ‘year’

[+ CANONICAL]	[-CANONICAL]
Criterion 2: Phonological distinctiveness	Criterion 1: Fused exponence (<i>partially</i>)
Criterion 4: More than one variant	Criterion 3: No overt realisation
Criterion 5: Morphological distribution	Criterion 10: No overlapping
Criterion 6: Alternation (overabundance)	Criterion 11: Absence of remainders
Criterion 7: Less relevant features involved	
Criterion 8: Contextual features	
Criterion 9: More than one feature	
Criterion 10a: Non-directional overlapping	
Criterion 12: Uniqueness	
Criterion 13: No outside conditions	
Criterion 14: Absence of syntactic effects	

6.5.2. The noun ‘person’

CRITERION 1: Fused exponence:

As with the noun ‘year’, this criterion is hard to apply. The stems *lud-* and *duf-* have a transparent internal structure, so they are non-canonically suppletive in this respect. But, when it comes to *tfolovik-*, its canonicity will depend on the decision we eventually make about its SECOND GEN PL (or GRADNM). In most, if not in all varieties (of the ones I have evidence of) the SECOND GEN PL of this particular noun can appear under the form *tfolovik*; i.e. with a form that is *phonologically syncretic* with the NOM SG. For the sake of consistency with the analysis for ‘year’, I analyse this form as heteroclitic (thus taking mark *zero*, which is characteristic of inflectional classes I and III). So in the same vein as with the noun ‘year’, the most sensible solution seems to say that there is no fused exponence; and hence, that according to this criterion the noun ‘person’ does not meet the canon either.

CRITERION 2: Phonological distinctiveness

All three stems are phonologically very different and they all have different etymologies. Thus, in this respect, there is no doubt that the noun ‘person’ meets the canon.

CRITERION 3: No overt realisation

The noun ‘person’ fails to meet this canonicity criterion, as the phonological realisation of the stem is mandatory in every cell of the paradigm.

CRITERION 4: More than one variant

For the majority of the idiolects recorded, the possibilities of suppletion are limited to the choice between two stems; i.e. either *tšovovik-* or *ljud-*. These varieties make the suppletive noun ‘person’ less canonical in this respect. But a very small number of idiolects in the corpus alternate between three stems: *tšovovik-*; *ljud-*; and *duf-*. Hence, the noun ‘person’ is closer to the canon than ‘year’ in this respect. Furthermore, this puts the West Polesian noun ‘person’ amongst the most canonically suppletive of the Slavonic varieties (together with Bulgarian and Macedonian).

CRITERION 5: Morphological distribution

At first glance, the distribution between *tšovovik-* and *ljud-* follows a morphosyntactic pattern (or morphosemantic distribution); that is to say, the alternation between one stem and another is clearly motivated (or easily predicted) by a (single) feature value: **NUMBER**. Thus, so far, according to this criterion, it is less canonical. Nevertheless, the presence of the stem *tšovovik-* in the SECOND GEN PL (OR GRADNM) (which, by the way, is virtually present in most speakers’ idiolects), makes it harder to describe with a single feature value. This way, the distribution seems to follow a morphological/morphomic pattern instead. In addition to this, when it comes to *duf-*, the analysis becomes more blurred (which aligns with the adnumerative form(s)), especially, if we disregard the two instances found in which it is not bound by a numeral. Thus, the noun ‘person’ has a more morphosyntactically motivated distribution than the noun ‘year’. Nevertheless, there is still evidence to support the idea that the noun ‘person’ meets this canonicity criterion.

CRITERION 6: Alternation

The GEN PL cell is particularly prone to overabundance, particularly in the cell of the SECOND GEN PL (or GRADNM). For example, speakers like Z4, T1 and Tor1 display different stems with higher numerals (see Table 35). It is far from the level of overabundance for ‘year’, yet, in this respect, the noun ‘year’ is still canonically suppletive.

CRITERION 7: Relevance of the features involved

The distribution or division between *tfolovik-* and *ljud-* encodes NUMBER (i.e. a relevant feature). So, in this respect, the noun is less canonical but still more canonical than if it were a verb. With the form *duf-*, it does not matter that much which analysis we have chosen for (CRITERION 5, *supra*) since the ADNUMERATIVE feature value is very specific and not very relevant (see also cross-linguistic evidence for this in (§5.1.2.)). Thus, those idiolects with the stem *duf-* have a more canonical type of suppletion, than those where it is not present. Yet, the form *tfolovik-* as a GEN PL (or GRADNM) exists in the vast majority of idiolects. This, being a combination of features or an irrelevant feature value, makes the noun ‘person’ more canonical in this respect.

CRITERION 8: Contextual features

The case of *duf-* is a good candidate for a contextual feature, given that the adnumerative cells are determined by the presence of numerals. The other two stems may help to mark NUMBER, but they do that on their own (i.e. without the need of a context), and yet, in the SECOND GEN PL (or GRADNM) cell, the stem alternation is not

marking **NUMBER** (or not at least in the way it is marked in the rest of the **PLURAL** sub-paradigm). Thus, in this respect, the noun ‘person’ meets the canonicity criterion.

CRITERION 9: More than one feature

If we are to treat the **ADNUMERATIVE(S)** as a **NUMBER** value, then none of the suppletive stems meets this canonicity criterion, since suppletion only marks **NUMBER** distinctions. But given that in the previous chapter we had concluded that the **ADNUMERATIVE** is not a **NUMBER** value (i.e. but a hybrid-feature value, instead), the stem *duf-* makes it more canonical in this respect, because it involves a combination of **CASE (DIRECT)** and **NUMBER (NON-SG)**.

CRITERION 10: No overlapping

At first glance, the fact that there is a remainder of the stem *tfolovik-* in the **PLURAL** sub-paradigm suggests an overlap (and thus, less canonicity). Nevertheless, the direction in which the overlap is happening is arguable.

Diachrony should not be decisive for measuring the canonicity of this criterion, although it must be said that is unclear. The first option is that as we have said for the noun ‘year’ (where the stem *fi**od* was an intrusion), the stem *lud-* was the original one. Thus, the stem *lud-* has been progressively overtaken by *tfolovik-* (and the Ukrainian [NOM SG] form *ljudina* would be an argument for this).

The second alternative is to state that the stem *lud-* has slowly overtaken the **PLURAL** sub-paradigm from *tfolovik-*. Since *tfolovik-* is more prominent (with the exception of

Ukrainian, it is the most extended form in Slavonic) and can appear in the SINGULAR as well as a part of the PLURAL sub-paradigm, I am more inclined to argue for this second option. In any case, the ambiguous directionality of the overlap makes it more canonical in this respect.

With respect to the stem *duf-*, it is a lot harder to determine whether it has been a later intrusion into the system, since it still exists with an independent meaning. Since the direction of the overlap is also unclear, it makes the noun ‘person’ more canonical in this respect.

CRITERION 11: Absence of remainders

Corbett (2007: 23) proposes the same example for Russian (i.e. its cognate) as an illustration of a suppletive noun paradigm with remainders. In (132) we see that, certainly, West Polesian replicates the same pattern in the GEN PL/GRADNM cell.²²⁰

(132) a. (T3.2.1 09:44)

poʃtʃita-u jix; ʃestnatsat tʃolo'vik na mene adn-o
count.PRF-PST.M.SG 3PLACC sixteen person.GRADNM to 1SG.ACC one-N.SG

‘I counted them: sixteen people to one, just me.’

b. (Tor1.25 05:31)

vosjim tʃolo'vik ma-l-i-sja rostrilje-ti.
eight.NOM/ACC person.GRADNM have-PST-PL-REFL shoot.PRF-INF

‘They were about to shoot eight people.’

²²⁰ Note that although (130) b. and (132) c. are very similar, they were produced at different times, and thus are different utterances.

c. (Tor1.25 07:29)

vin	tɪsatʃɪ	lu'd-ej	spas
3SG.NOM.M	thousand	person-GEN.PL	save.PRF.PST.M.SG

‘He saved a thousand people.’

Interestingly enough, (132) b. and c. show that the double options for marking the GEN PL/GRADNM can also be brought to the ACC PL cell with this noun. Yet, the existence of an alternative stem, *duf-*, in some idiolects makes the whole picture a lot more complex, in this respect. But, in general, it can be said that the noun ‘person’ fails to meet this canonicity criterion.

CRITERION 12: Uniqueness

On the one hand, the stem *duf-* cannot be derived further (or I have not found any instance of it in the corpus). On the other hand, I have found derived forms of *tʃolovik-* and *lud-* in the corpus.

(133) a. (Tor 1.12 02:11)

tak-ija	dobr-ija	'ljud-k-i
that-NOM.PL	good-NOM.PL	person-APPRECIATIVE-NOM.PL

‘Such nice (little) people.’

b. (TL3.3 00:07) [...]

s	'ljud-sk-ofio	pol-a,	s	'ljud-sk-ix
from	person-POSS-GEN.SG.M	field-GEN.SG	from	person-POSS-GEN.PL

saraj-uv,	postroi-ti	sobi	saraj-i
barn-GEN.PL	build.PERF-INF	REFL.DAT	barn-ACC.PL

‘They built themselves barn[s] out of people’s fields and barns.’

c. (TL6.2 01:28)

a	prixod-I-sja	jakisja	tʃolo'vi-tʃok
and	come.IMPF-3SG-REFL	some.NOM?	person-DIMINUTIVE.NOM.SG

‘And a little man/person’ arrives.’

That is to say *tʃolovitsʃok* and *ljudki* are repeating the formal correlation between *tʃolovik-* and *lud-*. Thus, in this respect, the noun ‘person’ does not meet the criterion to be canonically suppletive.

CRITERION 13: No outside conditions

As we have seen in the section above (§6.3.3.), there are hardly any conditions that trigger one stem or another (other than **NUMBER**). Thus, for this criterion, the noun ‘person’ is closer to the canon.

CRITERION 14: Absence of syntactic effects

The choice of one stem or another does not have any syntactic effects. It has no other function than to inform of a **CASE** or a **NUMBER** change, and this is still arguable, since the presence of remainders makes the morphosemantic distribution a lot less predictable. In any case, in this respect, the noun ‘person’ meets the canon.

Summary:

I have summarised the results of the canonicity test in Table 37:

Table 37 Summary of the canonicity criteria for the noun ‘person’

[+ CANONICAL]	[-CANONICAL]
Criterion 2: Phonological distinctiveness	Criterion 1: Fused exponence (<i>partially</i>)
Criterion 4: More than one variant	Criterion 3: No overt realisation
Criterion 5: Morphological distribution	Criterion 10: No overlapping
Criterion 6: Alternation (overabundance)	Criterion 11: Absence of remainders
Criterion 7: Less relevant features involved	Criterion 12: Uniqueness
Criterion 8: Contextual features	
Criterion 9: More than one feature	
Criterion 10a: Non-directional overlapping	
Criterion 13: No outside conditions	
Criterion 14: Absence of syntactic effects	

6.5.3. Overall summary

If we compare the noun ‘year’ to the noun ‘person’ in West Polesian, we can see that the noun ‘year’ (11/15) fulfils more canonicity criteria than ‘person’ (10/15).

Moreover, the tendency for ‘person’ is to have less alternating stems than ‘year’ and the distribution of stems is less marked by a morphosyntactic pattern than for the noun ‘person’. Thus, the noun ‘year’ is closer to the canonical instance of suppletion.

I have summarised the results in Table 38:

Table 38 Summary of the canonicity criteria for the nouns ‘year’ and ‘person’

	West Polesian ‘year’	West Polesian ‘person’
1. Fused exponence	–	–
2. Phonological distinctiveness	+	+
3. No overt realisation	–	–
4. More than one variant	+	+
5. Morphological distribution	+	+
6. Overabundance	+	+
7. Less relevant feature	+	+
8. Contextual feature	+	+
9. More than one feature	+	+
10. No overlapping	–	–
10a. Non-directional overlapping	+	+
11. No remainders	–	–
12. Unique	+	–
13. No outside condition	+	+
14. No syntactic effects	+	+

6.6. Summary and conclusions

I have used two specific lexical items in West Polesian, the nouns ‘year’ and ‘person’, in order to explore suppletion. I have shown that the stems that alternate in the paradigms are not mere synonyms, for there is cross-Slavonic evidence as well as within West Polesian of them being suppletive. Furthermore, even in the varieties where there is a two or three-way (form) paradigm, the distribution of the stems is still unequal: *ɦod-*, is undoubtedly more frequent than *rk-*; and the same applies to *tɦolovik-*, which is more prominent than *duf-* and *lud-*. I have shown that the conditions for predicting the speakers’ choice of one over another stem did not take us very far. I have then presented an analysis of these suppletive nouns using Corbett’s (2007) criteria. The conclusions extracted from this are that except for the presence of remainders, the lack of phonologically null or fused stems and the overlapping of paradigms, the West Polesian noun ‘year’ is very close to the canon (or the most extreme possible case of suppletion). The noun ‘person’ has a more common (and thus less canonical) behaviour, particularly because of its tendency to morphosyntactic distribution (patterning) and less stem alternation than ‘year’. Hence, data from West Polesian call for a careful revision of two main traditional assumptions in linguistics.

The first one is to do with suppletion itself. I have shown that there is no evidence to believe that in order for suppletion to happen the stems must be either ‘filling a gap’, or competing or specialising (what I called “The Armistice Theory”) or following any morphosemantic divide. The paradigms look more like an empty grid in which the speakers can insert one form or another with relative freedom without affecting the meaning.

The second one is to do with variation and overabundance. I have shown how the nouns ‘year’ and ‘person’ in West Polesian can display complex patterns of suppletion which vary from one speaker to another. It is commonly assumed that when more than one linguistic form (with apparent identical meaning) is possible, there must be a (hidden) sociolinguistic or discourse-level motivation behind the choice. In the 401 tokens of ‘YEAR’ from free speech utterances, I have not found any correlation between age or gender (not even location in many cases) and the choice of a particular stem. As for the noun ‘PERSON’, out of 245 tokens, the only partly meaningful factor has been age. I have not been able to identify the stem *duf-* in the youngest speakers’ speech (e.g. born after 1950) but that only suggests that it is in the process of extinction.

In any case, such an assumption would derive from a narrow understanding of the nature of language, which Dorian (2010) blames on the tradition of descriptive work carried out by linguists who speak a highly standardised language. However, in speech communities with a very horizontal structure and lacking a standard form of their language (such as East Sutherland Fisherfolk Gaelic, where Dorian conducted extensive fieldwork), it is not uncommon to find a high level of inter- and intra-speaker variation with no social weighting. This seems to be the case in Western Polesie as well: a population prominently composed of farmers (with the exception of people who leave the community for temporary jobs, mostly as builders), socially stigmatised because of their language and economic status, they form quite a homogeneous group. In this fashion, acknowledging that variation is possible without necessarily resulting in competition of forms is a plausible explanation to be considered as well.

Dorian (2010) emphasises that intra-speaker and inter-speaker variation with no social weighting (or as she calls it “sociolinguistically neutral variation”) is a lot more common than it may have been traditionally described. However,

“[l]inguists socialized in standard-language environments bring their own ideologies to their descriptive work, importantly including what James Milroy calls “the belief that a ‘language’ must exist in some authoritative, invariant form” (1999: 17)” (Dorian 2010: 287).

Although most West Polesian speakers I have interviewed (in Belarus) were aware that they did not speak (Standard) Belarusian, (Standard) Ukrainian or Polish (i.e. feeling ethnically and linguistically different from those national languages), they could hardly tell me the name of their language or variety (see (§1.2.1.)).

Thus, although these are very interesting instances of suppletion, we should not be very surprised to obtain similar results in other languages as awareness of the phenomenon increases amongst the community of linguists. Instead of always searching for complex explanations to give account of great variation (such as ‘obscure’ and ‘hidden’ discursive properties we are not aware of), I extend Dorian’s (2010) call to all linguists to embrace variation.²²¹

²²¹ “I emphasize the apparent ordinariness of Embo village to make the point that what might be deemed ‘exotic’ findings can appear in settings that do not appear the least bit exotic. If unexpected linguistic findings can emerge from a village in present-day Scotland, squarely within the industrialized Western world, it seems reasonable to suppose that there are other settings, some of them perhaps equally ordinary-seeming and some more obviously different from the settings most linguists are personally familiar with, that harbor equally unanticipated sociolinguistic phenomena” (Dorian 2010: 312).

Section II. Future tense

Chapter 7

West Polesian FUTURE TENSE

Dahl (2000b) studied the grammaticalisation of the FUTURE TENSE across European languages. He proposed different “families”, according to the morphological structure and the semantics of the constituents involved. A small number of languages, such as Finnish, lack any FUTURE TENSE constructions, whereas Romance languages have robust FUTURE TENSE constructions. Data from West Polesian suggest it has at least six constructions which correspond to the FUTURE TENSE families proposed by Dahl (2000b).

On the one hand, this should not be surprising, as, according to the cross-linguistic surveys in Bybee (1985) and Ultan (1978), FUTURE TENSE grams are reformed frequently. On the other hand, no European language appears to have more than three grams for the FUTURE TENSE in Dahl’s (2000b) survey, which would make West Polesian a ‘record holder’ within Slavonic languages, and most likely also in European languages. This represents an interesting situation from the point of view of typology, as well as historical linguistics, as this finding can assist in tracing areal influences.

First, I present the main FUTURE TENSE REFERENCE constructions I have documented in West Polesian and introduce some of the theoretical problems associated with the FUTURE TENSE in typology (§7.1.). Second, I examine each construction individually and describe their specific functions or semantic nuances (§7.2.). Third, I present a general analysis of the grammaticalisation of the potential FUTURE TENSE constructions (§7.3.). I apply several tests or criteria to determine the level of grammaticalisation

of each of the forms, thereby providing morphological evidence that proves that they are genuinely grammaticalised (i.e. inflectional) forms. Fourth, I look at those constructions from a historical and typological perspective (§7.4.), summarising current thinking about their etymologies and drawing parallels with other European languages. I show that the six constructions have some peculiarities, but that they can still be framed within the usual semantic bases for FUTURE TENSE constructions; and that many of them are indeed areal developments. Finally, I summarise the main results and draw general conclusions (§7.5.).

7.1. Introduction

In this section I introduce the main six FUTURE TENSE REFERENCE constructions that I have identified in West Polesian, which match the categories in Dahl's (2000b) survey (§7.1.1.). Then, I present the theoretical difficulties FUTURE TENSE constructions pose to typology (and general descriptive work) and outline the expectations for FUTURE TENSE forms cross-linguistically (§7.1.2.). I close this section by commenting on the specificities of the methodology used for obtaining data for this chapter (§7.1.3.).

7.1.1. The FUTURE in West Polesian

West Polesian displays a rich variety of morphosemantic strategies (or grams) to express futurity. Even looking at a fairly refined version of West Polesian, such as Klimčuk's translation of the New Testament (NT) (2010),²²² we can observe a rich variety of forms within a few verses).

²²² Klimčuk (2010) said in his foreword that his translation was based on other existing NT translations in Slavonic languages including Belarusian, OCS, Polish, Russian and Ukrainian.

Klimčuk's (2010) translation of the NT in WP²²³ (Matthew 5: 5-8)

- (134) ⁵ ʃtʃaslɪv-I [...] tɪx-I, bo jixn-ɪju **stan-e** **zɪmnj-a.**
happy-NOM.PL meek-NOM.PL as POSS.3PL.-INS.SG become.PRF-3SG earth-NOM.SG
- ⁶ ʃtʃaslɪvɪ ɦolodn-I i prafnuʃtʃ-I pravd-I, bo vonɪ **nasɪt-jatsjt-sja**
happy-NOM.PL hungry-PL and thirsty-PL truth-GEN.SG as 3PL.NOM fill.PRF-3PL-REFL
- ⁷ ʃtʃaslɪv-I mɪlostɪv-I, bo vonɪ **bud-utj** **pomɪlovan-I.**
happy-NOM.PL merciful-NOM.PL as 3PL.NOM be.FUT-3PL comforted-PL
- ⁸ ʃtʃaslɪv-I ʃtʃɪr-I serts-em, bo vonɪ **batʃɪ-tɪ-mutj** Boɦ-a.
happy-NOM.PL pure-NOM.PL heart-INS.SG as 3PL.NOM see-INF.HAVE.3PL God-ACC.SG

“⁵ Blessed are the meek, for the earth will become theirs. ⁶ Blessed are those who hunger and thirst for righteousness, for they will be filled. ⁷ Blessed are the merciful, for they will be shown mercy. ⁸ Blessed are the pure in heart for they will see God.”

The corpus I have gathered for this project is even richer in forms. Thus, by comparing these materials with Dahl's (2000b) survey of FUTURE TENSE constructions in European languages, West Polesian FUTURE TENSE constructions fall into at least six of his FUTURE TENSE construction-families:

1. The Ukrainian SYNTHETIC (IMPERFECTIVE) FUTURE (Dahl 2000b: 319). It consists of the infinitive form of the verb + a cliticised form of the verb ‘to have’ with PERSON/NUMBER marking.

²²³ English glosses based on (NIV, 2011). Translation modified to look more like the WP version.

West Polesian

(135) (Z10.1 00:20)

ja ʃ:jo ni **zna-ti-mu** skaza-ti
1SG.NOM Q.ACC.SG NEG know-INF-have.1SG say-INF

‘I won’t be able to tell [it] (lit. I will not know to tell).’

2. The DE-OBLIGATIVE FUTURE (2000b: 323).²²⁴ It consists of a verb originally meaning ‘have to’ (in this case in the REFLEXIVE form) + a verb in the infinitive.

(136) (T12.8.el 04:32)

vona **maj-it-sa** **rodı-ti** v ijulj-e
3.SG.F.NOM have-3SG-REF deliver-INF in July-LOC.SG

‘She is due to (deliver in) July.’

3. The Balkan DE-VOLITIVE FUTURE (2000b: 323).²²⁵ It consists of a verb originally meaning ‘to want’ + a verb in the infinitive.

(137) (Tor1.36.pr 01:51)

vona **xotf-e** kot-a **pokormi-ti**
3.SG.F.NOM want-3SG cat-ACC.SG feed.PERF-INF

‘She is going to feed the cat (lit. She’ll need to feed the cat).’

²²⁴ Dahl (2000b: 318-319) calls this the “Balkan ‘have’ future”, although this is not the only DE-OBLIGATIVE FUTURE-construction family in Europe. He previously mentions the “Romance inflectional future”, which is analogous to this form. However, in order to avoid confusion with the rest of the terms here (i.e. not to confuse it with the INFLECTIONAL or SYNTHETIC FUTURE, as in *robitımu* ‘I will do’), and for the sake of consistency, I use a more generic term ‘DE-OBLIGATIVE’.

²²⁵ Dahl (2000b: 322) mentions the existence of another DE-VOLITIVE construction-family in Northern Europe (Germanic). Yet, the West Polesian DE-VOLITIVE is genetically and typologically closer to the Balkan (i.e. Southern Slavonic) family than to the Northern European one.

4. The ‘BECOME’-TYPE or DE-VENITIVE FUTURE (2000b: 323-324). It consists of an auxiliary verb meaning ‘to become’ + a verb in infinitive.

(138) (Tor1.53.el 03:28)

jek	vona	stan-e	[xorofo]	futʃi-tr-sja,
COMP	3SG.NOM.F	become.PRF-3SG	well	study-INF-REFL

to	jiʃi	da-dutj	premiʃ-u
then	3SG.DAT.F	give.PRF-3PL	prize-ACC.SG

‘If/when she will start to study hard, she will get a reward/prize.’

5. The Slavonic PERFECTIVE FUTURE (or PRESENT) (2000b: 323, 326).²²⁶ It consists of a PERFECTIVE verbal base, with PERSON/NUMBER inflection (as in the PRESENT TENSE).

(139) (Z2.6 03:07)

batjuʃk-a,	jak	ʒe	ʒ	ja,	[fiet-ohio],	sober-u
priest-VOC	how	then	PART	1SG.NOM	this.GEN.SG	gather.PRF-1SG

fiet-ije	pir-ije?
this-ACC.PL	feather-ACC.PL

‘Father (Orthodox priest), how will I gather all those feathers?’

6. The Slavonic COPULAR CONSTRUCTIONS (2000b: 324), which consist of the auxiliary ‘to be’ in the FUTURE TENSE + a verb (in this case, in the infinitive).

²²⁶ Dahl (2000b) as well as other authors mentioned here use the form ‘Slavic’, which is far more widely used. However, since this thesis is due to be presented in Great Britain and for the sake of consistency, I have replaced every instance of ‘Slavic’ for ‘Slavonic’, which is considered the norm in British academia.

(140) (P2.5.pr 00:04)

a	ʃ:jo	mi	bud-em	robi-ti	na	Pasku?
so	Q-ACC.SG	1PL.NOM	be.FUT-1PL	do-INF	in	Easter.LOC.SG?

bud-em	praznova-ti
be.FUT-1PL	celebrate-INF

‘So what are we going to do for Easter? We are going to celebrate it.’

In addition to these, there may be a small number of other constructions (such as **pustitsj(a) + infinitive*). I deal with the ‘hapax constructions’ in greater detail later on (in (§7.2.7.)).

So how is it possible to have so many constructions or structures for expressing FUTURE TENSE in a Slavonic language or even in a European language? Are all these constructions in West Polesian fully interchangeable (i.e. absolute synonyms) or do they add different semantic flavours to the actions expressed? Are they all part of the ‘regular grammatical system’, also known as ‘inflectional’ in certain frameworks? I address all of these questions within this chapter.

So far Dahl’s (2000b) is the most complete survey of the different constructions, or morphosemantic strategies, across European languages, and thus I take his categories as a point of departure. For the sake of simplicity, I will keep part of Dahl’s (2000b) terminology when referring to the constructions under study.²²⁷

²²⁷ The exceptions to this are the terms ‘DE-OBLIGATIVE’ (see footnote 224) and ‘DE-VENITIVE’, which Dahl (2000b) mainly uses for FUTURE TENSE constructions with the verb ‘come’ as their basis. He refers to what is here termed ‘DE-VENITIVE’ as the “‘become’-type future”, which seems rather inelegant or non-technical. Moreover, because verbs meaning a change of state can also be DE-VENITIVE (from Latin *devenio*, meaning both ‘arrive to’ and ‘turn (in)to’ (Online Latin Dictionary)), and these are a far more

7.1.2. Theoretical problems about the FUTURE TENSE

“Tenses are puzzling to linguists, not to language users in a discourse” (Gvozdanović 1994: 199).

Among the most common three TENSES (i.e. PAST, PRESENT and FUTURE), the FUTURE is certainly the rarest and the biggest source of disagreement between linguists when discussing TENSE values.²²⁸ First, according to Dahl (2000b: 309), the FUTURE is epistemologically very different from the PRESENT and the PAST TENSES. “We cannot perceive or remember future states of affairs, and it has been disputed whether statements about the future can be said to have a determinant truth value” (Dahl 2000b: 309-310). For this reason, Dahl (2000b: 310) concluded that ontologically or semantically there are at least two archetypes of FUTURE TENSES in the world’s languages: “intention-based” FUTURES and “prediction-based” FUTURES.

Second, Slavonic languages have very restricted ways of marking futurity. Thus, finding indices of an inventory of constructions (six) that is not only too big for Slavonic, but also cross-linguistically, appears questionable. Nevertheless, according to Bybee (1985: 158), FUTURE TENSE forms are reformed frequently, suggesting that the FUTURE is the most open value of all the TENSES cross-linguistically. Bybee (1985: 158) highlights that this instability of FUTURE TENSE forms is related to “the incorporation in the future meaning of a modal and other notions which make meaning changes away from the temporal more likely”. Hence,

common FUTURE TENSE semantic bases than those simply denoting ‘to come’ (at least in European languages), I use the term ‘DE-VENITIVE’.

²²⁸ For example, according to Ultan (1978: 85), “[...] future tense markers [...] generally tend to be more marked than either present or past tenses”.

we should not be surprised by a larger inventory of constructions than for other TENSE values, because these constructions are an expected result of the semantic bases (usually modals) upon which they are typically built.

Third, according to Bybee (1985), the FUTURE TENSE tends to have a different morphophonological realisation from the PAST and the PRESENT TENSE cross-linguistically. According to her “[...] if futures develop independently, then their status as members of the *tense* category would be ambiguous, which it is” (Bybee 1985: 159). Consequently, there is an ongoing debate among Slavists on whether the FUTURE is really a TENSE value in Slavonic or not. Certain linguists look reluctantly on the legitimacy of COPULAR constructions, as in (140), as genuine FUTURE constructions in Slavonic languages (mostly Russian and Polish). The belief is that, since COPULAR forms do not comprise a single morphophonological unit, they are outside of the ‘main’ (i.e. synthetic) system, or that, if they belong to the system, they belong to a secondary system. Moreover, the idea of defining all the morphosyntactic features in binary values (e.g. Jakobson (1932)) has led some Slavists to even deny the existence of a FUTURE TENSE as a value at all in Slavonic (see more in Vater 1995: 162-163). This could be also said about most of the FUTURE TENSE constructions, except for the PERFECTIVE and the SYNTHETIC FUTURE, under study here. However, Mönke (1971) and Whaley (2000), among others, have argued against such a hypothesis.

Whaley (2000), based on Bybee’s (1985) and Ultan’s (1978) surveys, explains that FUTURE TENSE morphosemantic constructions, or grams, are very frequently realised periphrastically in the world’s languages. Thus, there is no reason to delegitimise the value of certain constructions under study here, or automatically assume they are a

calque from a non-Slavonic language, just because they are realised periphrastically. It is a ‘natural’ typological tendency. Therefore, I address the etymology of these West Polesian constructions in (§7.4.).²²⁹

An even stronger typological argument is that, in most Slavonic languages, PRESENT and FUTURE TENSE forms distinguish all PERSON and NUMBER values, whereas PAST TENSE forms only distinguish GENDER and NUMBER. Thus, the fact that the FUTURE TENSE is realised periphrastically is eclipsed by the use of a different verbal marking system for the PAST.

In summary, due to its particular morphological realisation and its ontological value, some people may be reluctant to call the future tense a tense value. However, we have seen that there are solid typological arguments to argue against such a position. The future is quite a dynamic form (historically), and has a transparent etymology (with often recognisable semantic bases that repeat cross-linguistically), which causes it to be periphrastic. Thus we should not be surprised to find different periphrastic future tense constructions in West Polesian, which may be at different levels of grammaticalisation and which may be interfering with each other. What is particularly interesting about West Polesian is that the inventory of these constructions is the biggest in the Slavonic family, gathering virtually every existing construction in the family, and that some of these constructions have acquired special uses and connotations, as I demonstrate in the following sections.

²²⁹ “Thus, in the case of Slavic, it might be best to view analytic future constructions not as unusual from the standpoint of the Slavic verbal system, but as typical from the standpoint of future-tense constructions in the world's languages. In sum, one cannot argue that a construction in Slavic is of foreign origin based on the fact that it is analytic” (Whaley 2000: 85).

7.1.3. Remarks about methodology

The future tense has been by far the hardest parameter to elicit. Whilst the material for the previous chapters has relied heavily on utterances retrieved from the corpus of transcribed free texts, this has not been possible for the future tense. Due to its delicate semantic nature (given that it is easily replaced by other forms) and relatively low frequency in the corpus, many of the examples provided here have been obtained through direct elicitation (most often using multiple choice questionnaires and grammaticality judgments, and, less frequently, using translations or rephrasing exercises). This is a common problem: Ultan (1978: 89) also pointed out the lower saliency of the future in his cross-linguistic survey of the future tense.²³⁰ I have also paid attention to overheard utterances outside the context of elicitation, as these often revealed forms that did not appear during the sessions. Finally, I have included a few examples from Fjodor Klimčuk's translated portions of the NT (2010), with the caveat that the West Polesian he used is a fairly standardised version of the language.

7.2. Uses and functions of the examined constructions

In this section I examine the morphological structure of the aforementioned constructions. Moreover, I describe the various uses they can have and the different nuances between them.

²³⁰ "Another index of future markedness is low frequency of occurrence vis-à-vis past and present tense. This factor has already been pointed out by Greenberg (1966: 87-88) for VEDIC SANSKRIT, LATIN and RUSSIAN [...]" (Ultan 1978: 89) (see references in original text).

7.2.1. The SYNTHETIC FUTURE

Most descriptions of Ukrainian, as well as the few on West Polesian and Southwestern Belarusian, say that the SYNTHETIC FUTURE is an identical synonym of the COPULAR FUTURE TENSE.

As I show in (§7.3.2.3.), for the vast majority of cases the SYNTHETIC and the COPULAR FUTURE TENSE share the same ASPECT restrictions, attaching exclusively to IMPERFECTIVE infinitives. Nevertheless, the use of the SYNTHETIC FUTURE is far less frequent, even in those varieties where all the FUTURE constructions analysed here are commonly used. The use of the SYNTHETIC FUTURE is generally more restricted with INTRANSITIVE verbs, and it is also very infrequent with REFLEXIVE verbs, most likely due to their phonological complexity.

7.2.1.1. Morphological structure

West Polesian SYNTHETIC FUTURE TENSE is composed by an (IMPERFECTIVE) infinitive + a suffix with an eroded form of the verb *mati* ‘have’ (in the PRESENT TENSE) which codifies the PERSON/NUMBER of the SUBJECT and TENSE.²³¹

- (141) (B21.ov) tʃaj **pi-ti-miʃ**?
 tea.ACC.SG drink-INF-HAVE.2SG
 ‘Are you having [lit. drinking] [any] tea?’

²³¹ Some may argue that the form is **imati*, but I deal with this discussion in (§7.4.1.).

Table 40 Synthetic future paradigm of the verb *robiti* ‘to do; to make’

1SG	ro'biti-mu	1PL	ro'biti-mo
2SG	ro'biti-meʃ/miʃ	2PL	ro'biti-mite/mita
3SG	ro'biti-me/ma	3PL	ro'biti-mut(j)

The infinitive *'piti* ‘to drink’ is the base that bears the main semantic load and the ASPECT; whereas *-miʃ* encodes TENSE and PERSON/NUMBER. In the case of REFLEXIVE verbs, the *-sj(a)* suffix is added after the PERSON/TENSE suffix, but it is rare.

(142) (Tor1.el)

zaxo-di	do	mena,	a	to	ja	obizati-mu-sj
enter-IMP.2SG	to	1SG.GEN	and	so	1SG.NOM	offend-HAVE.1PL-REFL

‘Come to my place, or I’ll be offended.’

During one of the sessions, Tor1’s son was observing the interview, and commented on forms like “*blinu bız masla 'sklejovatımutsj*” ‘the pancakes will get stuck without [any] butter’, which Tor1 had just accepted as correct. He said that, while correct, such forms are “too long for people to use them”, so they would use other forms, especially the COPULAR. This seems a reasonable observation, and judging by the instances of SYNTHETIC FUTURE forms in the corpus, the principle seems to apply to all the varieties under study, especially for those infinitives stressed on the third syllable from the right.

7.2.1.2. Uses and connotations

Mackevič (1959: 203-204) remarks that the SYNTHETIC FUTURE form in Southwestern Belarusian dialects, and thus, by extension, West Polesian, is used alongside the analytic form, but that **no differences** in meaning have been identified. However,

“[...] when two consequent or simultaneous actions in the future are presented, often both morphological forms are used [...]. Most frequently the closest action is expressed with the analytic future form and the subsequent with the synthetic form” (Mackevič 1959: 203-204) [My translation].

I have not been able to demonstrate such a tendency in my corpus of West Polesian but have instead determined that, apparently, the SYNTHETIC FUTURE construction is more suitable for questions, whereas the DE-OBLIGATIVE is more appropriate for general statements and predictions.

- (143) a. [Situation: T2 was talking about his daughter and her family who were supposed to visit them on that day, but it was already evening time and they had not appeared yet. T2 told me about them that “*Vonŭ malisja prŭjixatrŭ*” [DE-OBLIGATIVE] ‘They were going to come’. But five minutes later, he changed his mind and decided to ring his daughter to check out whether they were still coming on that day, so he asked the following.]

(T2.ov)

–alo! vita **prŭjixa-tŭ-mita?**
hello 2PL.NOM arrive.PRF-INF-HAVE.2PL

‘–Hello, are you going to come?’

- b. [Situation: It's very cold here during winter and that is why your friend's children take her home during the cold season. But this year it's going to be warm. How do you ask your friend whether she's staying for the winter or not?]

“If you are asking you say “*tʃi zabratmutj*?” [SYNTHETIC] ‘will [you] be picked up?, but if you are replying “*majutsja zabratr*” [DE-OBLIGATIVE], “*zabrutj*” [PERFECTIVE FUTURE] (‘they are going to pick [me] up’)” (T5.el 03:08) [My translation].

Having said this, there seems to be a reduced tendency for the SYNTHETIC FUTURE to be used for predictions and prophecies, over the COPULAR form.

- (144) a. (B4.1 00:59)

nu odn toj ʃtʃo na neb-i znaj-e,
so one.NOM.SG.M that NOM.SG.M RELNOM in heaven-LOC.SG know-3SG
to i **sudr-ti-me!**
so and judge-INF -HAVE.3SG

‘The one in heaven knows and will judge!’

- b. (T1.9 01:09)

silsk-i, silsk-ofio bere, bo ka.... florodsk-uju
villager-NOM.SG villager-ACC.SG take-SG as say[3PL?] urban-ACC.SG.F
vozm-e to una ni **umij-ti-me** u sil-i
take.PRF-3SG so 3.SG.NOM.F NEG know-INF-HAVE.3SG in village-LOC.SG
[...] vs-jo robi-ti
all-ACC.SG.N do-INF

‘People from the countryside take [as spouses] people from the countryside because they say that if [he] takes a woman from the city she won’t know how to [live] and do things in the village.’

Apparently, the SYNTHETIC FUTURE denotes less certainty than the COPULAR form, at least in contexts of semi-rhetorical questions.²³²

[Situation: We can't see Kolja. So we are wondering what he may be doing].

(145) a. ^I (Tor1.59.3.el 01:07)

ja duma-ju, na primjer, zare **mitſta-ju**,
1SG.NOM think-1SG in example.ACC.SG? now dream-1SG

ſ:jo zare Kolja **robi-ti-me?** ja **mitſta-ju**
Q.ACC now K.NOM.SG do-INF-HAVE.3SG 1SG.NOM dream-1.SG

ſ:jo vin zare **robi-ti-me**
Q.ACC 3SG.NOM now do-INF-have.3SG

a. ^{II} * ja mitſta-ju ſtſo Kolja zare **bud-e** **robi-ti**
1SG.NOM dream-1.SG Q.ACC K.NOM.SG now be.FUT-3SG do-INF

‘I think, I wonder (lit. ‘dream’) what will Kolja be doing now?’

b. (T6.5.pr 01:25)

ni znaju ſtſe [...] ſ:jo vin ſtſe **robi-ti-me.**
NEG know-1SG yet Q.ACC 3SG.NOM.M yet do-INF-have.3SG

‘I still don’t know what he is going to do.’

And I have found an instance where it may retain some obligative meaning.

²³² On occasion it can also be used for promises, as in (195) b. (*infra*).

- (146) (TL6.10 01:08) [Context: People were being deported to Germany. A boy had been called to go to an appointment, where he knew he was going to be cheated and deported with the rest. So he fled.]

vs-i **pr-jd-utj**, a vin rodɪ ɪ **itɪ-me**, vonɪ **pɪjd-utj**
 all-NOM.PL go.PRF-3PL so 3SG.NOM kind_of go-HAVE.3SG 3PL.NOM go.PRF-3PL
 ‘Everyone will go, and he apparently has to go as well, they will go [...].’

Besides these prospective uses, the SYNTHETIC FUTURE is also used to talk about events or situations taking place in the present, which are the outcome of something that would have been unexpected in the past.

- (147) a. (B6.11 09:07) [...]

u odn-ej xat-ɪ ʒɪ-l-ɪ, trɪ nɪvɪstki; tɪpɐ na
 in one-LOC.SG house.LOC.SG live-PST-PL three.NOM girl.ADNM now in
 dvor-ɪ **nɪ ʒɪ-tɪ-me** i dvi nɪvɪstki, nɪ trɪ
 property-LOC.SGNEG live-INF-HAVE.3SG and two.NOM girl.ADNM NEGthree.NOM

‘[...] three single girls used to live in [the same] house. Nowadays not even two or three girls would [lit. will] live in the house.’

- b. (T1.8 00:25)

fɪeta Kolxoz zroβɪ-l vot tut daβna, i
 this.NOM.SG.F Kolkhoz.NOM.SG do.PRF-PST.M so here long_ago and
 vid-ɪʃ, zam-ɪɐr-z, nɪ-ma komu **topɪ-tɪ-mutj**.
 see-2SG freeze.PRF-PST-M.SG NEG-have.PRS Q.DAT.SG.M heat-INF-HAVE.3PL

‘The Kolkhoz [collective farm] made one long time ago, but you see, it’s gone, they don’t have people for whom they will light [the banja] up.’

7.2.2. DE-OBLIGATIVE FUTURE

The ‘have’-type or DE-OBLIGATIVE FUTURE TENSE construction has a strong sense of **intentionality** (in some cases it even retains some obligational flavour). After asking some speakers about the meaning of *majusj robıtı* ‘I will do’ (DE-OBLIGATIVE), some (e.g. T2) said that it has the same meaning as *budu robıtı* (COPULAR), and thus that there is no difference between the two. Yet after several weeks some speakers (T5, T10 and T12 i.a.) pointed out that the auxiliary *majusj* of the DE-OBLIGATIVE is the equivalent of the Russian *sobirajus’* ‘I intend to’.

7.2.2.1. Morphological structure

The DE-OBLIGATIVE FUTURE construction in West Polesian is composed by the auxiliary verb *matı* ‘to have’ (at least synchronically) + REFLEXIVE suffix + infinitive. The auxiliary bears PERSON/NUMBER and probably also TENSE information, whereas the infinitive encodes the main semantic load and ASPECT.

- (148) ja ma-ju-sj (po)sadı-tı kartopl-i
 1SG.NOM have-1SG-REFL plant.PRF-INF potato-ACC.PL
 ‘I am going to plant potatoes.’

7.2.2.2. Uses and connotations

In most instances the DE-OBLIGATIVE FUTURE is inappropriate for general predictions. However, in situations such as the discussion of someone’s pregnancy, this form

appears the most natural to use, so I suspect there is also some lexicalisation behind the choice of one structure over another.²³³

(T9, T12.el)

(149) a.^I vona **ma-et-sa** **rodi-ti** v ijunje [DE-OBLIGATIVE]
 3SG.NOM.F have-3SG-REFL deliver-INF in June-LOC.SG

a.^{II} *vona **rodi-ti-me** v ijun-je [SYNTHETIC]
 3SG.NOM.F deliver-HAVE.3SG in June-LOC.SG
 ‘She is due (to deliver) in June.’

b.^I vona **ma-et-sa** **rodi-ti** dotʃk-u [DE-OBLIGATIVE]
 3SG.NOM.F have-3SG-REFL deliver-INF daughter-ACC.SG

b.^{II} vona **rodi-ti-me** dotʃk-u [SYNTHETIC]
 3SG.NOM.F deliver-INF-HAVE.3SG daughter-ACC.SG
 ‘She will give birth to a girl.’

Because of its morphological structure, the auxiliary *matu* can appear in the PAST TENSE and indicate prospectivity in the PAST (neither the COPULAR, nor the PERFECTIVE or the SYNTHETIC FUTURES allow this), for example, as in (132) b. in the previous chapter.

(132) b. (Tor1.25 05:31)

vosjim tʃolo'vik **ma-l-i-sja** **rostrilje-ti.**
 eight.NOM/ACC person.GRADNM have-PST-PL-REFL shoot.PRF-INF

‘They were about to shoot eight people.’

²³³ About (149) b^I- b^{II}. The SYNTHETIC form is generally good for both general predictions and intentions. In the context of discussing pregnancy and birth *roditime* implies a greater certainty than the DE-OBLIGATIVE form *maetsa roditi*, according to my consultants. Given that people can expect, but not know for sure when they are going to give birth, the synthetic form sounded inappropriate to all three of my consultants. However, when it comes to sex of the baby, T13 argued that there can be more certainty about it (thanks to technology), thus, it is okay to speculate (using the de-obligative) about the sex of the baby, but it is also acceptable to assert it quite firmly (if we know the results from the doctor).

According to T6, the use of this form implies a high level of intentionality, with a sense of commitment “*namjereno, vže zadumano*” ‘well-thought, already planned’,²³⁴ whereas T9 emphasised that the DE-OBLIGATIVE expresses more uncertainty than the PERFECTIVE:

“So if it’s tomorrow, I will say “tomorrow *pojīdu* ‘I will [already] go [PERFECTIVE]’ to Drahičyn, or Brest, or Pinsk... I will go tomorrow”. But if I say *majusja jixatī* ‘I’m due to go [DE-OBLIGATIVE]’ it is unknown when *pojīdu* ‘I will go [PERFECTIVE]’ (T9.9.el 02:32) [My translation].

The auxiliary *majusj* can be used on its own as a **free or unbound auxiliary**, like the COPULAR auxiliary *budu*. I do not have many examples of this in the corpus, but one instance is particularly worthy of comment. TL1’s daughter was supposed to water the potatoes one day, but she did not. While she was explaining to her mum why she had not been able to water them, (TL1.ov) asked “*a maišsja?*”, meaning ‘but are you going to water the potatoes?’

According to HL1’s comments, there is a correlation between ANIMACY and the morphological strategy used. I proposed (150) a., where the object is a HUMAN (children), and she objected to it saying that “you would say *maju*, if you had an item, but with children, you say *treba*”.

- (150) a. (HL1.el)
- | | | | |
|----------------------|--------------|------------------|-----------|
| * ma-ju-(sja) | dit-ej | položi-tī | spa-tī |
| have-1PL-REFL | child-GEN.PL | lay.PRF-INF | sleep-INF |
- (HL1.vol) > > minjī treba dit-ej položi-tī spa-tī
- 1SG.DAT necessary child-GEN.PL lay.PRF-INF sleep-INF
- ‘I need to/will put the children to sleep’.

²³⁴ (T6.3 01:57).

- b. praz pjet fiód-ov **maj-ut-sja** **dostroi-ti** dorofi-u
in five year-GEN.PL have-3PL-REFL build.PRF-INF road-ACC.SG
- (HL1.vol) > > sobiraj-ut-sja dostroi-ti; objazan-i dostroi-ti
plan-3PL-REFL build.PRF-INF obliged-NOM.PL build.PRF-INF
- ‘In five years time they will have to have finished building the road.’

7.2.3. The DE-VOLITIVE FUTURE

Some of the varieties surveyed in the West (of Western Polesie) have a DE-VOLITIVE FUTURE. In these varieties the verb *xotiti* ‘to want’ has a deontic meaning, and it seems that *majusj* (DE-OBLIGATIVE) is close in meaning to this, as it was the first response I obtained whenever I tried eliciting sentences containing *maju/ majusj*.

7.2.3.1. Morphological structure

The DE-VOLITIVE FUTURE TENSE is composed by an inflected form of the verb *xotiti* + infinitive. Most commonly the verb *xotiti* appears inflected in the PRESENT, which means that the topic-time is in the FUTURE; but less frequently the verb can appear in the PAST TENSE if the topic-time is in the PAST (i.e. FUTURE-IN-THE-PAST).²³⁵ As with other periphrastic forms, the auxiliary encodes PERSON, NUMBER and probably also TENSE information, whereas the infinitive bears the biggest semantic load (of the verb). In

²³⁵ The FUTURE-IN-THE-PAST is less clear for the DE-VOLITIVE. I have not been able to prove whether there are any TENSE restrictions, but I have a strong suspicion that there may be (unfortunately my main consultant, Tor1, passed away during my research so I have not been able to confirm this). I have certainly identified utterances with the verb *xotiti* in the PAST TENSE + INFINITIVE in the corpus, although in most of them there were signs that it only had a ‘plain’ volitive meaning (i.e. no temporal or prospective value).

theory, both IMPERFECTIVE and PERFECTIVE infinitives can appear with *xotiti*. In practice, when I suggested periphrases with *xotiti* (among other constructions) to combine with an infinitive to my main informant, Tor1, she rarely picked the IMPERFECTIVE infinitives with *xotiti*. On the occasions when she did, she also chose both the IMPERFECTIVE and the corresponding PERFECTIVE infinitive.

(151) a. (Tor1.el)

zavtra	xotſ-u	jixa-ti	u	boljnits-u,	bo
tomorrow	want-1SG	go-INF	to	hospital-ACC.SG	as
treba	operatsij-u	zrobi-ti			
need	operation-ACC.SG	do-INF			

‘Tomorrow I will have to go to the hospital to have an operation.’

b. ja **xotſ-u** **zaruba-ti** pivnj-a, bo vin
 1SG.NOM want-1SG kill.PRF-INF rooster-ACC.SG as 3SG.NOM.M

v	mena	klu-je	i	može	i	popad-e	do	dit-aj
in	1SG.ACC	bite-3SG	and	may	and	attack.PRF-2SG	to	child-GEN.PL

‘I will need to kill the rooster as he keeps biting and he may attack children.’

As elsewhere in Eastern Slavonic, there is also a verb *xotiti*² which means ‘to want’. It is hard to determine whether it is a completely different entry at this stage, but it is very clear that the DE-VOLITIVE FUTURE (synchronically with a deontic value) originated from it.

(152) (Z10.12 00:30)

a	tipera	nikto	ni	xotſ-e	za	firoſ-i	do	tibe	i-ti
and	now	nobody.NOM.SG	NEG	want-3SG	for	money-ACC.PL	to	2SG.GEN	go-INF

‘So nowadays nobody wants to come [to help] you [even] for money.’

7.2.3.2. Uses and connotations

The DE-VOLITIVE FUTURE has very restricted uses as a FUTURE TENSE periphrasis. The verb *xotiti* denotes a type of obligation that is not (at least directly) imposed by external forces, such as the weather, or an institution like a bank.²³⁶ If the verb *xotiti* is used with an externally imposed obligation it loses its deontic meaning and just has (or preserves) a volitive meaning, which as a consequence sounds unacceptable to the speakers.

- (153) a. (Tor1.el) [Context: A man has been outside the house for a while. He really has to get into the house quietly.]

vin	xotʃ-e	tixenko	pri-ti	bo	dʒiti	sp-jetj
3SG.NOM	want-3SG	silently	enter-INF	as	child.NOM.PL	sleep-3PL

‘He will need to get in silently because children are sleeping.’

- b. (Tor1.el) [Context: There hasn't been any rain for a while.]

*ja xotʃ-u poli-ti ofiorod, bo vs-jo posoxn-e
1SG.NOM want-1SG water-INF garden.ACC.SG as all-NOM.SG.N dry.PRF-3SG

(Tor1.vol) > > treba	(obizatilno)	poli-ti	oñorod-a
be necessary	compulsorily	water-INF	garden.GEN?SG

'I (will) need to water the garden or else everything will dry.'

Thus, the DE-VOLITIVE form is closer to intentions than to predictions, but it still fails to fully adjust to any of these categories.

²³⁶ Otherwise other forms are used; e.g. *ja kraj maju* (T6), *ja dolžm, munji musovo* or *munji treba*. Tor1 was also aware of this restriction when I tried eliciting different forms (Tor1.45) “*ja xotʃu, fiato, ʃ:jo to takoje, znatʃit tʃastni, ʃ:jo ja [no] xotʃu; fiato nado, treba platiti*” ‘*ja xotʃu* [I have to] is something private, but [with your bills] you must pay [them]’.

7.2.4. The DE-VENTIVE FUTURE

Some of my language assistants from Aljaksiejavičy/Sičyv pointed out that *stanu robiti* (in the FUTURE) is not a West Polesian form, but Russian, although they recognised its existence for the PAST TENSE. Nevertheless, I overheard the form in informal conversations enough times that I could not ignore it, and there is some evidence in the corpus as well.

7.2.4.1. Morphological structure

With a similar structure to the rest of the periphrastic forms under study in this work, the DE-VENTIVE FUTURE is formed by a finite form of the auxiliary verb **statı*, marking TENSE and NUMBER/PERSON/GENDER, + infinitive, which bears the main lexical/semantic load. As opposed to the DE-VOLITIVE form, the DE-VENTIVE is far more common and acceptable for speakers in the PAST TENSE than in the FUTURE.

(154) a. (Z7.1 00:58)

a fiet-ije fiadı v miʃk-u **sta-l-i** ʃipi-tı,
so this-NOM.PL beast.NOM.PL in sack-LOC.SG become-PST-PL whisper-INF

sta-l-i **voruʃi-tı-sa!**
become-PST-PL shake-INF-REFL

‘So the beasts in the sack started whispering and shaking!’

b. (Z1.7.1 07:46)

ja vʒe ni zna-ju kilke fiód-ov vʒe
1SG.NOM already NEG know-1SG how_many year-GEN.PL already

ni **sta-l-i** **bra-tı** jiʃio
NEG become-PST-PL take-INF 3SG.ACC.M

‘I no longer know how many years ago they stopped taking him.’

Nevertheless, both uses are possible; i.e. the auxiliary appears in the PERFECTIVE FUTURE if the topic-time is in the FUTURE, and most frequently **statr* appears in the PAST TENSE if the topic-time is in the PAST.

(155) a.^I (Tor1.el)

ja zautra **stan-u** **robi-ti** u Drofiŋin-i
1SG.NOM tomorrow become.PRF-1SG.FUT work-INF in Drahičyn-LOC.SG

a.^{II} (Tor1.vol) > > ja zautra uže **stan-u-sj**,
1SG.NOM tomorrow already become.PRF-1SG.FUT-REFL

stro-ju-sj na robot-u
enrol-1SG.FUT-REFL in work-ACC.SG

‘I will start working in Drahičyn tomorrow.’

b. (Tor1.el)

ja sta-u robi-ti na fjet-aj robot-i,
1SG.NOM become-PST.PL work-INF in that-LOC.SG work-LOC.SG

no minji ni ponravi-l-o-sj
but 1SG.DAT NEG like.PRF-PST-N.SG-REFL

‘I started working in that job, but I didn’t like it.’

There are at least two verbs with the infinitive **statr*: one is an auxiliary; the other one is a polysemous verb. Alongside the different lexical entries or possible meanings of the verb *statr* (and variations), the specificity of the form, which Whaley (2000) calls *stat*² for its cognate in Russian, is also highlighted by differences at the level of morphology.

The first example, the (auxiliary) verb **statr*¹, is a defective verb only available for the PERFECTIVE and, as a consequence, it lacks a PRESENT TENSE (as with the COPULAR form, synchronically; i.e. *budu*). It also lacks an infinitive; in fact, I tried eliciting an

infinitive form from several speakers, but they were unable to produce one. I have no clear evidence to make a strong claim about West Polesian, but the cognate of **stat* in Russian (*stat'*, which elsewhere else behaves like West Polesian) in its second meaning “has an imperfective counterpart [...] *stanovit'sja*, that has the meaning ‘become’” (Whaley 2000: 141).

7.2.4.2. Uses and connotations

Most often the DE-VENITIVE FUTURE is closer to an INCHOATIVE periphrasis than to a purely grammaticalised or semantically neutral FUTURE TENSE construction. Due to these specific semantic constraints, this was the construction more often rejected in the elicitation sessions. That is to say, besides all the additional meanings (or homophonous forms), the DE-VENITIVE (the first entry) is more often a synonym of *potfati/natfati* (the proper verb for ‘to start’). Compare (156) a-b.

(156) a. (B6.44 03:20) [Obscure meaning]

natʃa-l-i kopa-ti, to mi duma-l-i ʃtʃo voni ni **vikopa-jutj**,
begin-PST-PL dig-INF so 1PL-NOM think-PST-PL COMP 3PL.NOM NEG dig.PRF-3PL

jak voni **vikop[o]jut** kanav-i, aʒ ni usva-ju
COMP 3PL.NOM dig.PRF-3PL channel-ACC.PL PART NEG assimilate.PRF-1SG

‘They started digging out the channels, we thought that they wouldn’t finish digging them as they were digging them, I couldn’t believe it.’

b. (B6.44 00:14)

a **sta-l-i** uže po rusko-mu **fiovori-ti**, jak poprijeza-l-i
so become-PST-PL already in Russian-DAT.SG speak-INF COMP arrive.PRF-PST-PL

z zarobitkov. Da uže [...] tam mi **bud-em fiovori-ti** to i to
from jobs-GEN.PL PART already there 1PL.NOM be.FUT-1PL speak-INF so and so

‘So we started speaking Russian when we came back from our temporary jobs [abroad]. And so we said we would say/speak this and that.’

The second entry for the verb **statr*² (i.e. when it is not an auxiliary verb) can mean ‘to become’:

(157) a. (B19.1 01:14)

potom dvatsat fiod, jak uže kroxi **stan-e**
then twenty year.GRADNM COMP already bit become.PRF-3SG

vzrozlif-im

mature-INS.SG

‘Then [he turns] twenty, and he becomes a bit [more] mature.’

b. (B20.21 00:50)

stan-eſ busjk-om bud-eſ ʒi-ti kaj tʃolovik-a
become.PRF-2SG stork-INS.SG be.FUT-2SG live-INF by human-GEN.SG

i bud-eſ lovi-ti fiad-ov

and be.FUT-2SG hunt-INF beast-ACC.PL

‘You will turn into a stork and you will live close to people, and you will hunt for insects.’

When used in **negative** sentences, it means ‘to disappear’.

(158) a. (B7.1 01:48)[...]

dalej jix i ni sta-l-o t-ij firof-I
later 3PL.GEN and NEG become.PST-N.SG that-GEN.SG? money-GEN.SG?

na t-ij knizts-I
in that-LOC.SG book-LOC.SG

‘So, then that money disappeared from the book.’

b. (Z7.7 01:55)

zlik-l-a, propa-l-a, ni sta-l-o jiji
fade.PRF-PST-F.SG vanish.PRF-PST-F.SG NEG become.PST-N.SG 3SG.GEN.F

‘She faded out, she vanished, she disappeared.’

Besides the two possible entries for the verb **statı*, it is important to remark that there is a homophonous verb (at least in some of the PERFECTIVE forms) meaning ‘to stand’, which can be misleading. The verb *stavatı/stojetı* is a regular (or full) verb. Compare (159) a. (‘stand’) and (159) b. as an aux. (DE-VENITIVE) and also meaning ‘become’.

(159) a. (T8.1 07:45)

to pivinj sta-v na lodk-u de i... [...]sto’j-itj
so rooster.NOM.SG stand.PRF-PST.M.SG in ship-ACC.SG where and stand-3SG
‘The rooster stood on the ship and [...] here it’s standing.’

b. (HL2.1 00:36)

vot prijixa-v vin do xat-i, sta-v¹ vin
so arrive-PST.M.SG 3SG.NOM.M to house-GEN.SG become-PST.M.SG 3SG.NOM.M

ljudj-am propovedova-tı, sta-l-o² po dv-a,
person-DAT.PL preach-INF become-PST-N.SG about two-NOM.M

po tri t[olo¹vik-a vjeru[t]f-ix [...]
about three-NOM person-GEN.SG? believer-GEN.PL

‘He arrived in the house and started preaching to people, two or three people converted.’

7.2.5. The PERFECTIVE FUTURE (*or* PRESENT)

7.2.5.1. Morphological structure

The PERFECTIVE FUTURE (*or* PRESENT) is composed of a PERFECTIVE verbal base + PERSON/NUMBER fused inflection realised as a suffix (160).

- | | | | |
|----------|---------------|----|---------------------|
| (160) a. | rob-lju | b. | zrob-lju |
| | do.IPFV-1SG | | do.PRF-1SG |
| | ‘I am doing.’ | | ‘I will have done.’ |

We could say that the suffix also codifies TENSE. However, this is not as easy to demonstrate as it may appear at first glance, because the PERFECTIVE PERSON/NUMBER endings do not always match those of the PRESENT TENSE, i.e. when the PERFECTIVE form is not derived by mere suffixation, as in most Slavonic languages.

7.2.5.2. Uses and connotations

Even though the PERFECTIVE FUTURE is generally used as a default form of FUTURE for PERFECTIVE verbal bases, it can still have some additional connotations, other than the ones given by the ‘perfectivising prefixes’.

Due to its PERFECTIVE ASPECT, it denotes a **strong certainty** that the predicated action will be fulfilled in the future.

- (161) a. (TL6.7 01:48)
- | | | | | | | |
|--|---------|-------------------|----------|----|---------|------------------|
| kudi | voni | namjeren-i | viz-ti, | to | voni | poviz-utj |
| where | 3PL.NOM | intended-PL | take-INF | so | 3PL.NOM | take.PRF-3PL |
| ‘Wherever they are intended to take [him], they will.’ | | | | | | |

b. (TL6.12 01:33)

I ja riſi-l-a ſtʃo zamuʒ ni poj-du
and 1SG.NOM decide.PRF-PST-F.SG COMP in_marriage NEG go.PRF-1SG

‘And I decided that I wouldn’t get married (lit. ‘I will not get married’).’

The PERFECTIVE FUTURE can be also used in predictions, prophecies, morals and the like:

(162) a. (B20.11 01:50)

jak krad-eſ, to ni potovstij-eſ
if steal-2SG so NEG fatten.PRF-2SG

‘If you steal, you won’t put on weight.’

b. (Z.4.2 04:29)

duma-l-i ona bud-e kolo svo-jej mat-eri,
think-PST-PL 3SG.NOM.F be.FUT-3SG around own-GEN.SG.F mother-GEN.SG

da moʒi kroxi otjam-et-sa, prijd-e do rozum-a
so maybe bit relax.PRF-3SG-REFL come.PRF-3SG to sense-GEN.SG

‘They thought that by being next to her mother she may calm down and come back to her senses.’

The PERFECTIVE FUTURE is also frequently employed in **orders**.

(163) a. (B20.10 00:24)

priſ-ov starost-a i kaza-u “vsj-o,
come.PRF-PST.M.SG boss-NOM.SG and say.PRF-PST.M.SG all.NOM.SG.N

prijd-eſ u rekrut-i”.
go.PRF-2SG to rekrut-NOM.PL?

‘The boss came and said “right, you’ll become a rekrut”.’

b. (B19.3.0 00:11)

ī batjk-o kaʒ-e [...] vetʃorom **prjɔd-ete** i
and father-NOM.SG say-3SG tonight come.PRF-2PL and

skaz-ete mnji koʃo vi batʃi-l-i [...]
say.PRF-2PL 1SG.DAT Q.ACC 2PL.NOM see-PST-PL

‘And the father said [...] “tonight you will get home and tell me whom you saw [...]”.’

7.2.6. The COPULAR FUTURE

7.2.6.1. Morphological structure

The COPULAR FUTURE is composed of the PPERFECTIVE FUTURE form of the verb *buti* ‘to be’ (acting as an auxiliary) + infinitive (IMPERFECTIVE). The auxiliary marks TENSE on its base, and PERSON/NUMBER inflection as a suffix.²³⁷ The non-finite form, the IMPERFECTIVE infinitive, bears the main semantic load or lexical content, as well as ASPECT marking. It is important to remember that, as in other Eastern (and most Western) Slavonic languages, the use of PERFECTIVE infinitives with this construction is ungrammatical.

(164) a. bud-u **robi-ti**
 be.FUT-1SG do.IPFV-INF
 ‘I am going to do.’

b. *bud-u **zrobi-ti**
 be.FUT-1SG do.PRF-INF
 ‘I will have done.’

²³⁷ Whilst for the regular PERFECTIVE FUTURE I have said that it is unclear which morpheme (if there is any at all) marks TENSE, when it comes to the form of *buti* ‘to be’, the auxiliary has a suppletive form for the FUTURE, so the whole form marks it.

7.2.6.2. Uses and connotations

The COPULAR FUTURE TENSE is by far the most frequent of all the constructions in the corpus. It is well-evidenced in every local variety in which I have worked. Moreover, when eliciting other FUTURE TENSE constructions, speakers replied most frequently with the COPULAR as a default form, making the task more difficult. Because of its semantic ‘neutrality’, it can be employed for most contexts requiring an IMPERFECTIVE FUTURE. The COPULAR FUTURE TENSE can be used for predictions as well as intentions.

- (165) a. (B20.7 00:47) [A curse pronounced by a wizard]

ja to jid-u, ale tɪ bilʃ **sidi-tɪ** tut nɪ **bud-eʃ**
1SG.NOM so go-1SG but 2SG.NOM more sit-INF here NEG be.FUT-2SG

‘I am riding [my horse], but you will not sit here anymore.’

- b. (TL6.3 02:11) [A prophecy]

v jak-uju por-u viʃixa-la z dom-u
in which-ACC.SG? time-ACC.SG? leave.PRF-F.SG from home-IIGEN.SG

ʊ tak-uju i bud-iʃ
in that-ACC.SG? and be.FUT-2SG

‘You will [return] home on the same day as when you left.’

The COPULAR form can also be used for instructions (166) a. and general truths (166) b.:

(166) a. (B20.2 00:28)

djid kupi-u jim bitjetsk-a, ka3-e
man.NOM.SG buy-PST.M.SG 3PL.DAT young_ox-ACC.SG say-3SG

“**bud-ete pasvi-tr; fionl-tr i pasvi-tr**”
be.FUT-2PL graze-INF drive-INF and graze-INF

‘The man bought them a young ox and said “you will graze it, you will drive and graze it”.’

b. (T11.20 01:40) [as a general rule]

smitan-a **bud-e** kisl-a **stoje-tr**
cream-NOM.SG be.FUT-3SG sour-F.SG stand-INF

‘The cream will become sour.’

7.2.7. Other marginal or ‘hapax’ constructions

7.2.7.1. *pustitisj(a) + infinitive

Among the rarely documented constructions, this is the most intriguing one typologically, as it does not resemble any of the descriptions provided by Dahl (2000b). It has only been utilised three times by two speakers (TL1 and TL4) from Tolково (Drahičyn), both of whom, unfortunately, I was not able to interview again.

One of the biggest challenges has been the lack of sufficient tokens of this construction in the corpus. This form has only been documented in the PAST TENSE, as a FUTURE-IN-THE-PAST. Moreover, there are clues that suggest this form would fail at least half of the grammaticalisation tests that I apply in (§7.3.) (e.g. TENSE-ASPECT-MOOD restrictions; semantic neutrality and generalisation).

Its semantics are vaguely reminiscent of the English expressions *let us (do (it))* > *let's (do (it))*. It is morphologically composed of the REFLEXIVE form of the verb *pustiti* (which literally means ‘to let; to release’, as in (167) a. + infinitive. Compare (167) b-d. with (167) a.:

(167) a. (Z2.2.1 04:10)

a vin ka3-e “**pust-i,** **pust-i,** **pust-i...**
 and 3SG.NOM.M say-3SG leave-IMP.2SG leave-IMP.2SG leave-IMP.2SG
 potomuʃjo i mene opotop-ɪʃ!
 because and 1SG.ACC sink.PRF-2SG

‘And he says: “release [me], release [me], release [me]... because you'll sink me as well!”.’

b. (TL4.1 02:40)

tam mnoʃa 3ɪnɪx-ɪv bu-l-o, tak-ɪx kavaljer-uv,
 there many suitor-GEN.PL be-PST-N.SG that-GEN.PL bachelor-GEN.PL
 to voni za nami jak **spusti-l-i-sja** **bɪʃ-tɪ [...]**
 so 3PL.NOM after 1PL.INS COMP release-PST-PL-REFL run-INF

‘There were so many suitors, and they were ready to chase after us as soon as we would start running [...].’

c. (TL1.8.1 02:08) [...]

kob vin xotj na ʃafi ot minja viʃ-ov,
 COMP 3SG.NOM just in step.ACC.SG from 1SG.GEN leave-PST.M.SG
 to ja v3e **pusti-l-a-sja** **vtika-tɪ**
 so 1SG.NOM already release-PST-F.SG-REFL escape.PRF-INF

‘[...] so that he would take one step further from me, I was already going to/ready to run away.’

d. (TL1.8.1 02:28)

no ja дума-л-а kob vin mene xotj na ʃafi
but 1SG.NOM think-PST-F.SG COMP 3SG.NOM 1SG.ACC just to step.ACC.SG

otpusti-v, ʃ:jo ja id-u za xljeb-om, i
release.PRF-PST.M.SG COMP 1SG.NOM go-1SG for bread-INS.SG and

ja vʒe tut **pusti-l-a-sja** **putika-ti**
1SG.NOM already here release-PST-F.SG-REFL escape.PRF-INF

‘I was thinking that he would step away from me at least a bit (lit. a step),
so I [could] bring bread, but I was going to/preparing to run away then.’

7.2.7.2. Pseudo DE-ANDATIVE constructions

DE-ANDATIVE FUTURE constructions (e.g. *I am **going to** sing*; *je **vais** chanter*) are well known in European languages. A few expressions denoting prospection which contained a motion verb appeared in the corpus of West Polesian. Taking into account the typological ‘popularity’ of DE-ANDATIVE FUTURE constructions amongst European languages, it is tempting to include the expressions found in West Polesian within the same group.

The two tokens I have identified in the corpus have in common that they use a motion verb (in both cases in the PAST TENSE) as an auxiliary and another motion verb as the main verb/non-finite form (168).

(168) a. (Z1.6.1. 03:18)

mi [ʃeto...] **uʒe** **ruxa-l-i** do xat-i **jixa-ti**
1PL.NOM this already move-PST-PL to house-GEN.SG go-INF

‘We were already planning to go home.’

b. (T2.9.1 00:31)

poslja v3a mi **poſ-l-i** v Subot **xod-i-ti**,
after already 1PL.NOM go.PRF-PST-PL to Subaty.GEN.PL? go.IMPf-INF

v3a tut vs-jo **prikrati-l-o-sj**
already here all-N.SG end.PRF-PST-N.SG-REFL

‘After that we started going to Subaty, everything here was over.’

I should remind the reader that these forms were not part of my core of research. Moreover, they have been found in isolated utterances, so I lack enough information to draw meaningful conclusions, other than documenting them and noting that more research is needed.

7.3. Grammaticalisation

Up to this point, I have identified six main FUTURE TENSE constructions, which resemble some of the categories or “families” that Dahl (2000b) had described. We have seen that all six constructions are used to express prospection or futurity in West Polesian. However, does it automatically make them grammaticalised (or inflectional) forms? In this section (§7.3.) I provide evidence from a morphophonological perspective arguing that all these six forms are genuine or grammaticalised FUTURE forms. I begin by defining grammaticalisation and presenting the most widespread understandings of this (§7.3.1.). I then run different tests to determine the degree of grammaticalisation of each of those constructions (§7.3.2.); and I close the section with a summary of the results (§7.3.3.).

7.3.1. Definition

Before proceeding further, let us define ‘grammaticalisation’. Dahl (2000c: 8-9) considers different conceptions of the phenomenon. The most widespread is to see grammaticalisation as “semantic bleaching, that is, the semantic content of the item is partly or wholly lost.” Other views emphasise the “death” of metaphors in a broader sense; obligatoriness; or the most classic Meilletian (1912) approach, which focuses on the diachronic process of semantic erosion. Dahl concludes the following about grammaticalisation:

“[i]n many cases, the crucial property is [...] that the use of a certain item is governed by factors other than carrying new and relevant information in the utterance context [...]” (Dahl 2000c: 8-9).

I will not go into much detail about the different approaches to grammaticalisation. Instead, I submit the constructions under study to different tests, based on properties commonly accepted to be part of ‘regular’ or ‘fully grammaticalised’ parts of speech, and assess the progress or position of each construction along the path of grammaticalisation.

7.3.2. Grammaticalisation tests

I have looked at the level of grammaticalisation of each of the structures suspected of being ‘proper’ FUTURE constructions. The following criteria have been used:

- Desegmentability; i.e. whether the constituents, if it is a periphrastic construction, allow the insertion of other constituents within them.
- PERSON, TAM (TENSE-ASPECT-MOOD) and VALENCE (transitivity) restrictions.

- Evidence of phonetic erosion.
- Evidence of additional meanings, which may uncover the etymology.
- Recursivity; i.e. whether the verb used for the base of the auxiliary can have the same verb as a complement.
- Ontology or type of FUTURE; i.e. whether they can be used to express intentions, predictions or both.

Some of these conditions or test criteria have been extracted from the EUROPTYP Questionnaire (Dahl 2000a).

7.3.2.1. Desegmentability

All the periphrastic FUTURE TENSE constructions allow the insertion of other constituents between the auxiliary and the infinitive.

(Extracted from free texts)

(169) a. (Tor1.38 01:12)

vo, i uoni vze prijaxa-l-i tudı, i vze
so and 3PL.NOM already arrive.PRF-PST-PL there and already

ma-l-i-sja nas **rostrılje-tı**
have-PST-PL-REFL 1PL.ACC shoot.PRF?-INF

‘So they came to us, and they were already about to shoot us.’

b. (T5.3 03:36)

xutko **bud-e** tsarov-a dotʃk-a zamuʒ_vixodı-tı,
soon be.FUT-3SG queen.POSS-NOM.SG.F daughter-NOM.SG get_married-INF

i **bud-e** objevlenj-e vıʃa-tı, ʒımx-i vıbıra-tı.
and be.FUT-3SG ad-NOM.SG hang-INF groom-ACC.PL? choose-INF

‘Soon the king’s daughter will get married and they will announce it
[as] they will be choosing grooms.’

c. (TL4.3 01:59)

i tudɪ **sta-l-ɪ** dom **stroj-tɪ**, **sta-l-ɪ**
and there become.PRF-PST-PL house.ACC.SG build-INF become.PRF-PST-PL

ljud-ɪ **kaj-tɪ-sj** **prixodɪ-tɪ** **sta-l-ɪ**,
person-NOM.PL repent-INF-REFL come.IPFV-INF become.PRF-PST-PL

ibo jistino bu-l-ɪ vjeruʃtɕ-ɪ
because truly be-PST-PL believer-NOM.PL

‘They started to build the house, people started to convert [lit. ‘repent’]
and to attend, because they were true believers.’

(Tor1.53.2,3. el))

(170) a. tapera ja **maj-u-sj(a)** pizdno **loʒɪ-tɪ-sj**
today 1SG.NOM have-1SG-REFL late lay-INF-REFL
‘Today I will go to bed late.’

b. **stan-u** ʋ Minsk-e **robɪ-tɪ**
become.PRS.PRF-1SG in Minsk-LOC.SG work-INF
‘I will start working in Minsk.’

c. vin **xotɕ-e** tixenko **zaj-tɪ** na xat-u
3SG.NOM want.PRES-3SG quietly enter-INF in house-ACC.SG

ʃtɕo dɪtɪn-a spi-tj
COMP baby-NOM.SG sleep-3SG
‘He will need to enter the house quietly because the baby is sleeping’.

It is worth noting that, with the SYNTHETIC FUTURE, we would not expect to be able to segment the constituents. However, when REFLEXIVE verbs are introduced, the internal syntax of the verb may be altered.²³⁸ As far as the West Polesian varieties I have

²³⁸ Mackevič (1959: 203) explains that in Southwestern Belarusian dialects which have the SYNTHETIC FUTURE, usually the REFLEXIVE suffixes (realised as *-sja*; *-sa*, or *-cca*) do not appear “after the infinitive, but after the inflection mark for person” [My translation].

surveyed are concerned, speakers tend to avoid the SYNTHETIC FUTURE construction with REFLEXIVE verbs (especially when they have more than two syllables), as I have already explained in (§7.2.1.1.).

In summary, all the periphrastic constructions allow the insertion of other constituents between the auxiliary and the main verb.

7.3.2.2. PERSON restrictions

Although more data are needed, there is evidence to suggest that all the FUTURE TENSE constructions are available for all PERSON values. Some PERSONS (more specifically 1SG, 3SG and 3PL) are more frequent in the corpus than others, especially when speaking about intentions. This is true both across West Polesian, and cross-linguistically. In light of the current data, PERSON does not affect the choice of construction.

(171) a. (Z.4.1.1 03:34)

mi	ponja-l-i,	ŝto	mi	vs-i
1PL.NOM	understand.PRF-PST-PL	COMP	1PL.NOM	all-NOM.PL
umr-em	fiolodn-aj	smjert-ju		
die.PRF-1PL	hungry-INS.SG	death-INS.SG		

‘We understood that we were all going to die of hunger.’

b. (Z2.1 04:38)

no,	mat-i,	fiet-u	podrobno	uže	3	ni
but	mother-VOC.SG	this-ACC.SG	thoroughly	already	PART	NEG
bud-em	rozkaz-uva-tr!					
be.FUR-1PL	tell-IPFV-INF					

‘But, mum, we will not tell such details!’

c. (Z10.10a 00:55)

xotʃ? To rozpi'ʃ-im-sja, da bud-em ʒi-ti²³⁹
want.2SG so register.PRF-1PL-REFL shall be.FUT-1PL live-INF

‘Do you want to? Let’s register [our union] and let’s live [together].’

Although impersonal sentences are very rare, I have documented a few instances where some of the constructions under study appeared. Therefore, there is no evidence to assume any restrictions for impersonal sentences.

For example, T14 was talking about the weather, on the first day of June. She had heard somewhere that if the first two days of June are rainy, then the rest of the summer will be dry. She therefore concluded with this sentence (172):

(172) (T14.ov)

jek i'di-ti-me, to bud-e sux-aj²⁴⁰
if go-INF-HAVE.3SG so be.FUT-3SG dry-INS.SG

‘If it rains (lit. ‘if it will go’), it will be dry.’

7.3.2.3. ASPECT restrictions

This is the biggest constraint with FUTURE constructions. As in every other Slavonic language, by definition, the PERFECTIVE FUTURE (or PRESENT) is only available for PERFECTIVE verbal bases. It is composed of a PERFECTIVE verbal base + PERSON

²³⁹ In this contexts the verb **xotʃ** is not being used as a future tense auxiliary (but in its primary volitive acceptance). I have not been able to find any DE-VOLITIVE contruction in which the auxiliary is used in a PERSON other than 1SG, 3SG or 3PL.

²⁴⁰ T8, T14’s mother says **itme**, so I am not sure whether I have properly recorded that form.

inflection. Consequently, using an IMPERFECTIVE verbal base + PERSON inflection unavoidably transforms the verb into a PRESENT TENSE form.

In contrast, the COPULAR construction is exclusively available for IMPERFECTIVE verbs, as is the rule in the majority of Slavonic languages (§7.4.6.). The SYNTHETIC FUTURE TENSE tends to replicate the behaviour of the COPULAR construction in terms of ASPECT.²⁴¹ Nevertheless, I have been able to document isolated instances where the morphological structure has extended to PERFECTIVE verbal bases, but they have very specific semantics. When I proposed that pattern (i.e. PERFECTIVE infinitive + *mu/mŭs/...*) to different speakers, the vast majority of forms were rejected, although there were some exceptions. Compare (173) a. with (173) b-c.

(173) a. (B8.4 03:09)

toɦdiʃi kaza-l-i [...] jak rodi-ti; to eto bud-e
then say-PST-PL if deliver-INF so this.NOM.SG be.FUT-3SG

xuzej boli-ti jak [x]to **uzna-ti-me.**
worse hurt-INF if someone.NOM.SG know.PRF-INF-have.3SG

‘Then, people used to say that giving birth is going to be more painful if anybody finds it out [lit. will know].’

b. (B6.20 06:12)

mi jak_to znaj-em "ve'silje", a jak kto,
1PL.NOM somehow know-1PL v.-ACC.SG? but if Q.NOM.SG

to ni **zna-ti-me**
so NEG know-INF-have.3SG

‘We know [what] "ve'silje" (‘wedding’) [means], but someone [else] won’t know it.’

²⁴¹ As has also been described for Standard Ukrainian and Southern Belarusian (Danylenko 2011, Mackevič 1959, Rusanovskij et al. 1986).

c. (B6.21.el)

- naŋotovi-l-a-sja na Paŋk-u!
 cook.PRF-PST-F.SG-REFL in Easter-LOC.SG?
- niŋs, zaŋtra **dojeda-ti-mo!**
 nothing tomorrow eat.PRF-INF-HAVE.1PL
- ‘- I have cooked too much [food] for Easter!’
- No worries, we’ll finish eating it tomorrow!’

West Polesian is innovative by extending the use of the SYNTHETIC FUTURE form to the PERFECTIVE. Interestingly enough, the oldest speakers (> 85 in 2017) are the ones who had used or accepted SYNTHETIC FUTURE TENSES with PERFECTIVE infinitives. Thus, it is not clear whether the extension of the SYNTHETIC FUTURE to the PERFECTIVE was once an active process, which for unknown reasons reverted at some point; or whether there is an ongoing change in West Polesian grammar that I have simply not been able to see in younger speakers’ speech. Amongst all the possibilities offered to Tor1 in the examples (174), only the DE-VENITIVE form with an IMPERFECTIVE infinitive was rejected.

(174) (Tor1.59.1.el)

vnuk-am	tiŋke	dva	fioɖi,	jek	rost-utj,
grandkid-DAT.PL	only	two.NOM.M	year.ADNM	COMPL	grow.PRF-3PL
ja_____	djiŋ-jum	konfet-i			
1SG.NOM	[BUY]	children-DAT.PL	sweet-ACC.PL		
<i>pokupaŋi/ kupljeti</i>	[IMPERFECTIVE]		<i>kupiti</i>	[PERFECTIVE] ‘to buy’	
budu pokupaŋi	[COPULAR]		majuŋj kupiti	[DE-OBLIGATIVE]	
*stanu kupljeti	[DE-VENITIVE]		stanu kupiti	[DE-VENITIVE]	
?pokupaŋimu	[SYNTHETIC]		kupitimu	[SYNTHETIC]	
			xotʃu kupiti	[DE-VOLITIVE]	

‘My grandkids are only 2, but when they grow older, I will buy them sweets.’

The DE-VENITIVE FUTURE has been a difficult construction to elicit in general. Most of the instances proposed were rejected, partly because the semantics were also problematic. Thus, I lack enough evidence to say whether any ASPECT restrictions apply to the DE-VENITIVE form. The *DABM* (1963 vol. I: 166) suggested the DE-OBLIGATIVE structure is a synonym of the COPULAR FUTURE and the SYNTHETIC FUTURE. However, evidence from the corpus and the interviews suggest that this is far from being true, at least aspectually. In light of the data, the DE-OBLIGATIVE construction and the DE-VOLITIVE, which is becoming more of a modal verb than a FUTURE construction, are the only versatile constructions in terms of ASPECT. In the examples in (175) a-b. we can see that they are combined with a PERFECTIVE infinitive (which seems the most natural combination, particularly, for the DE-OBLIGATIVE).

(175) a. (T5.5 07:22)

vona zara mij skaza-l-a, ʃto **zarɪza-tɪ** **maj-it-sja**
 3SG.F.NOM now 1SG.DAT tell.PRF-PST-F.SG COMP kill.PRF-INF have-3SG-REFL
 ‘She just told me that she is going to slaughter me.’

b. (T5.5 08:15)

tɪ [v] ʒ fiolov-u mene **zarɪza-tɪ** **xotʃ?**
 2SG.NOM in from head-GEN.SG 1SG.ACC kill.PRF-INF want.2SG
 ‘Are you going to behead me? (lit. ‘do you want to cut my head?’)

Moreover, in examples (176) and (177), we can see that both constructions are in theory available for both ASPECTS, but they tend to prefer the PERFECTIVE. Note that when the language assistant expressed a preference for one of the forms in a pair, I have underlined the preferred form.

(176) *batʃiti* /*po'batʃiti* (*brata*) ‘to see ([one's own] brother).’ (el. Tor 1.55)

<i>'batʃiti</i> [IMPERFECTIVE]	<i>po'batʃiti</i> [PERFECTIVE]
<i>budu 'batʃiti</i> [COPULAR]	<i>'budu po'batʃiti</i> [COPULAR]
<i>'batʃitimu</i> [SYNTHETIC]	<i>po'batʃitimu</i> [SYNTHETIC]
<i>#xotʃu 'batʃiti</i> [DE-VOLITIVE]	<i>xotʃu po'batʃiti</i> [DE-VOLITIVE]
<i>majusja 'batʃiti</i> [DE-OBLIGATIVE]	<i>majusja po'batʃiti</i> [DE-OBLIGATIVE]
	<i>po'batʃu</i> [PERFECTIVE FUT]

(177) *ru'batı* /*poru'batı* (*drova*) ‘to chop (wood)’ (Tor1.55.el)

<i>ru'batı</i> [IMPERFECTIVE]	<i>poru'batı</i> [PERFECTIVE]
<i>bude rubatı</i> [COPULAR]	<i>*budu porubatı</i> [COPULAR]
<i>rubatime</i> [SYNTHETIC]	<i>#porubatime</i> ²⁴² [SYNTHETIC]
<i>?xotʃe rubatı</i> [DE-VOLITIVE]	<i>xotʃe porubatı</i> [DE-VOLITIVE]
<i>majetsa rubatı</i> [DE-OBLIGATIVE]	<i>majetsa porubatı</i> [DE-OBLIGATIVE]
	<i>porubaje</i> [PERFECTIVE FUT.]

Besides, motion verbs also present some ASPECT restrictions for most speakers' varieties. For example the verb *iti* ‘to go [IPFV]’ was very often rejected when proposed with the COPULAR or the SYNTHETIC FUTURE + IMPERFECTIVE INFINITIVE, in constructions like **budu iti v firod* ‘I am going to go (on foot) to the city’.²⁴³ However, when motion

²⁴² T15 rejected this form.

²⁴³ Nevertheless, I have found exceptions in the corpus; e.g.:

a. (TL1.4.2 00:16) *tam dorofia i papo bude jixatı i z dorofı sjuda*
‘There’s a road and dad is going to go (by means of transport), and from the road up to here.’

verbs, such as *iti* ‘to go (on foot)’ are employed not in their primary sense (i.e. moving from one place to another), but in idiomatic expressions (e.g. talking about the weather), this appears more acceptable for most speakers.

(178) a. (T8.ov)

i-ti-me doʃtʃ
go-INF-HAVE.3SG rain.NOM.SG

‘It’s going to rain.’

b. (TL4.10 07:11)

nje, bab-o, ja ʃtʃe dva fioda
no granny-VOC.SG 1SG.NOM more two.NOM.M year.ADNM

zamuž ni **bud-u** **vixodi-ti!**
behind_husband NEG be.FUT-1SG exit.IPFV-INF

‘No, granny, I will not get married (lit. exit behind the husband) for another two years!’

c. (TL1.8.2 00:12)

ja skaza-u zavjedujuʃtʃ-im, ʃ:jo mi biʃ
1SG.NOM say.PRF-PST.M.SG director-DAT.SG? COMP 1PL.NOM more

ni **bud-em** **vixodi-ti**
NEG be.FUT-1PL exit.IPFV-INF

‘I told the director that we would not go [to work] outside anymore.’

b) (TL1.1 01:02) [no i...] riʃili v direvnju poiti, skritisja “može... tut front **bude idti**”
‘So they decided to go to the village, to hide “maybe... the front will go [through] here”.’

c) (T5.2.2 02:45) de jakomu takomu priʃeldʃevi, niʃ:jomu jak **bude idti** darofiju i bude prositi
“podarite!”, možna podariti.
‘You can give this to a vagabond, a beggar, when he will be walking down the road and begging
“give me!”.’

7.3.2.4. Restrictions according to TRANSITIVITY

All the constructions studied here are available for TRANSITIVE, DITRANSITIVE and INTRANSITIVE verbs. The only exception is the DE-VENITIVE FUTURE. Most of the proposed instances were rejected, but again, probably due to semantic restrictions. It is not surprising that speakers rejected sequences involving COPULATIVE and other INTRANSITIVE verbs, like **stanu buti* ‘(lit.) I will become to be’.

(179) (Elicited from T12, T3 and Tor1)

Intransitives

- a. *vin*_____ *v* *fiorod-i*
 3SG.NOM.M [WORK (INTR)]²⁴⁴ in city-LOC.SG
majitsa robiti [DE-OBLIGATIVE] **stane robiti* [DE-VENITIVE]
robitime [SYNTHETIC]
bude robiti [COPULAR]

‘He is going to work in the city.’

- b. *vona*_____ *v* *ijulj-e*
 3SG.NOM.F GIVE_BIRTH-(INTR) in July-LOC.SG
majitsa roditi [DE-OBLIGATIVE]
roditime [SYNTHETIC]
#bude roditi [COPULAR]²⁴⁵

‘She is due (to deliver) in July.’

²⁴⁴ NB: The verb *robiti* ‘to work’ does not present any problem for the DE-OBLIGATIVE (and DE-VOLITIVE) forms in IMPERFECTIVE. However, when *robiti* is used with its second meaning ‘to do (something)’ (and thus TRANSITIVE) it is preferred in PERFECTIVE.

²⁴⁵ T13 argued her rejection “because the delivery could happen before, who knows?” By this, we could infer that for at least T12 (and T13, who was listening and approving) the copular future denotes a higher degree of certainty than the DE-OBLIGATIVE, but the reality in the corpus seems a lot more blurred. In fact, she produced a counterexample to this slightly afterwards.

c. Verbs taking a prepositional phrase

ja_____do sɨʃtr-i
1SG.NOM [CALL] to sister-GEN.SG

majusj zvonitɨ [DE-OBLIGATIVE] *stanu zvonitɨ [DE-VENTIVE]
pozvonju [PERFECTIVE FUTURE]
budu zvonitɨ [COPULAR]
zvonitimu [SYNTHETIC]

‘I will call my sister.’

Transitives

d. ja_____xat-u
1SG.NOM [BUY] house-ACC.SG

kupljetimu [SYNTHETIC] *stanu kupitɨ/ kupljetɨ [DE-VENTIVE]
budu kupljetɨ [COPULAR]
majusj kupitɨ/ kupljetɨ [DE-OBLIGATIVE]

‘I am going to buy a house.’

For DITRANSITIVES see examples (174).

7.3.2.5. MOOD restrictions

The OPTATIVE (or SUBJUNCTIVE) does not have a dedicated morphology (i.e. it is not part of the TAM system). It is formed by the heading particle (*nu*)xaj followed by a verb in the FUTURE. All the constructions are available in the INDICATIVE sentences, but only some of the constructions are available for OPTATIVE sentences. On the one hand, only forms with the PERFECTIVE FUTURE have been documented in the corpus.²⁴⁶

²⁴⁶ This includes *butɨ* in the PERFECTIVE FUTURE, which is the base of the auxiliary in the COPULAR construction.

(180) a. (B20.2 03:49)

xaj **ʒe** **mene** **ɾɪʒ-ut(j)!**
 let PART 1SG.ACC slaughter.PRF-3PL

‘Let them slaughter me!’

b. (TL4.1 10:32)

do **t-ej** **bab-I** **ni** **pɪjd-em,** **xaj** **bab-a**
 to that.GEN.SG.F lady-GEN.SG NEG go.PRF-1PL let lady-NOM.SG

t-aja **ber-et** **t-oe** **sjen-o**
 that-NOM.SG.F take-3SG that-ACC.SG.N hay-ACC.SG

‘They said, “we won’t go to that lady’s, let her take that hay!”.’

c. (B18.3 01:09)

xaj **tebe** **xotj** **u** **fɔrʃk-u** **svar-it,** **a**
 let 2SG.ACC even in bag-ACC.SG? suffocate.PRF-3SG but

mij **serts-e** **ne** **rv-I!**
 POSS.1SG.NOM.M heart-ACC.SG NEG tear-IMP.2SG

‘Let him even suffocate you in a bag, but don’t [ever] break my heart!’

On the other hand, when elicited directly, the DE-OBLIGATIVE form was always rejected (and there are no traces of it in the corpus of naturally produced texts), perhaps due to the retention of some modal connotations. The SYNTHETIC FUTURE TENSE was also rejected in all the instances I proposed and, in the same vein, it does not appear in the corpus. However, it is not as ‘categorically’ clear as for the DE-OBLIGATIVE. Equally, there is no evidence of the DE-VOLITIVE construction, i.e. as a FUTURE TENSE auxiliary,²⁴⁷ with the SUBJUNCTIVE, and it is unclear whether this is possible. The forms with the

²⁴⁷ I.e. not in its primary volitive meaning (‘want to’).

DE-VENITIVE FUTURE were rejected, but, once again, this may be due to other reasons, such as the semantics of the auxiliary itself. Yet, it does not seem improbable that DE-VENITIVE constructions could be available for the SUBJUNCTIVE.

(181) (T9, T12 ,Tor1.el)

- a. xaj mnifio djitok **rod-jatj** [PERFECTIVE FUT.]
 let many child.GEN.PL deliver.PRF-3PL

*xaj voni **rodi-ti-mutj** mnifia dit-aj [SYNTHETIC]
 let 3PL.NOM deliver-INF-HAVE.3PL many child-GEN.PL

(T12.vol) > > xaj mnifio djitok **uvod-jatj** [PERFECTIVE FUT.]
 let many child.GEN.PL bring.PRF-3PL

‘May they have many children.’

- b. [Context: ‘My neighbour is going to repent/convert’]

mij susid **ma-ji-tsa pokai-ti-sa**
 POSS.1SG.M.SG neighbour.NOM.SG have-3SG-REFL repent.PRF-INF-REFL

nixaj **pokajitsa** [PERFECTIVE FUTURE]

nixaj **bude kartsa** [COPULAR]

*(ni)xaj vin **maritsa pokartisa** [DE-OBLIGATIVE]

*(ni)xaj **kartime** [SYNTHETIC]

‘May he repent!’

- c. [Context: ‘My son is planning to move to the village.’]

mij sin **maj-it-sa piri'jixa-ti** u sel-o
 POSS.1SG.M.SG son.NOM.SG have-3SG-REFL move.PRF-INF to village-ACC.SG

*nixaj **majitsja piri'jixati** [DE-OBLIGATIVE]

(T12.vol) > > xaj vin **prijide** [PERFECTIVE FUTURE]

‘May he move!’.

7.3.2.6. CONDITIONAL/TEMPORAL restrictions

West Polesian, like Belarusian and Ukrainian, does not distinguish CONDITIONAL clauses from pure TEMPORALS, at least traditionally.²⁴⁸ The conjunction or complementiser *jak/jek*, or less commonly *koli*, can be translated as either ‘when’ or ‘if’, and has cognates in Belarusian *jak*, *kali* and Ukrainian *jak*, *koly* (as in (139)). Following Tor1’s judgments, all the studied constructions can appear in the protasis of a CONDITIONAL/TEMPORAL clause.²⁴⁹

(Tor 1, T12.el)

(182) a.ⁱ *koli vin ma-et-sja ustava-ti, to xaj minji pozvon-it.*
 when 3SG.NOM have-3SG-REFL wake-INF so may 1SG.ACC ring.PRF-3SG

a.ⁱⁱ *koli vin ustava-ti-me, to xaj minji pozvon-it.*
 when 3SG.NOM wake-INF-HAVE.3SG so may 1SG.ACC ring.PRF-3SG

‘When (or ‘if’) he wakes up, he should (lit. ‘may he’) call me.’

b. *koli vin stan-e xorošo zaroblje-ti, to xaj*
 when 3SG.NOM become.PRF-3SG well earn-INF so may
vin kup-it mam-e xat-u.
 3SG.NOM buy.PRF-3SG mum-DAT.SG house-ACC.SG

‘When (or ‘if’) he starts making a good wage, he should (lit. ‘may he’) buy his mum a house.’

²⁴⁸ I have documented instances of *jesli*, which is a conditional complementiser. It this seems more the result of language contact than a strategy in the language (i.e. it is probably a loan from Russian). Neither BLM, nor ULM distinguish between TEMPORAL and CONDITIONAL clauses, and they do not have any word like *jesli* (although Polish does).

²⁴⁹ I have also found an instance of the PERFECTIVE FUTURE used in a protasis in a free text:

(B20.11 01:50) *jak kradęś, to ni potovstijeś*

‘If you steal, you won’t put on weight.’

- c. *koli vin xotf-a jixa-ti na operatsi-ju, to jofio*
 if 3SG.NOM.M want-3SG go-INF to operation-ACC.SG so 3SG.ACC.M

zabe'r-a skor-a[ja pomož]
 pick.PRF-3SG ambulance-NOM.SG

‘If (or ‘when’) he will need to get an operation, the ambulance will pick him up.’

- d. *jak maj-i-sja prijixa-ti, to pozvo'n-i, d-aj zna-ti*
 if have-2SG-REFL arrive.PRF-INF so call-2SG.IMP give.IMP.2SG know-INF

‘If you are to come, let [us] know.’

7.3.2.7. Event time restrictions

Neither the COPULAR FUTURE, nor the PERFECTIVE FUTURE, nor the SYNTHETIC FUTURE can be employed in the PAST TENSE due to morphological restrictions. Nonetheless, this does not mean that they cannot be found inserted in a subordinate clause in which the main verb is in the PAST TENSE as in (185) c. On the contrary, the DE-VENITIVE, the DE-VOLITIVE and the DE-OBLIGATIVE constructions can appear in the PAST TENSE as finite forms (in a main clause) to make reference to events in the PAST with prospection.

- (183) a. (Examples from texts)

(Tor1.25 02:29)

to pap-u sudi-l-i za fiet-o, zabra-l-i,
 so dad-ACC.SG judge-PST-PL for this-NOM.SG take.PRF-PST-PL

arestova-l-i... i vže ma-l-i-sja sudi-ti
 arrest-PST-PL and already have-PST-PL-REFL judge-INF

‘They sentenced Dad for that, they took him, they arrested him... and they were already about to sentence him [to death].’

- | | | | | |
|---------|-------------------|------------------|------------|---------|
| v3e | sta-l-i | kuplja-ti | SVOJ-I | v3e |
| already | become.PRF-PST-PL | buy-INF | own-ACC.PL | already |

(Tor1.53.2.el)

- | | |
|-------------------|--------------|
| operatsij-u | (bo musovo) |
| operation- ACC.SG | as necessary |

b. *naʃ-i susid-i sta-l-i kupitɪ dom
1PL.POSS-NOM.PL neighbour-NOM.PL become-PST-PL buy-INF house.ACC.SG

[vol. Tor1] > > **xotʃ-utj** / **xoti-l-i** kupi-ti dom
 want-3PL want-PST-PL buy-INF house.ACC.SG

c. vin sta-u robi-ti tam bo duma-u ʃtʃo __ mnifia
3SG.NOM.M become-PST.M.SG work-INF there as think-PST.M.SG COMP a lot

polutjatime	[SYNTHETIC]
stane zarobljetı	[DE-VENTITIVE]
bude zarobljetı	[COPULAR]
> > [vol. Tor1] zarobıtj	[PERFECTIVE FUT.]

‘He started working [there], because he thought he was going to make a lot.’

383

had overheard some people in the village, including his grandmother, using it. I will expand more on the use of DE-VENITIVES in Eastern Slavonic in (§7.4.4.).

Notwithstanding the fact that event time restrictions apply to some constructions, the choice of the construction is unrelated to remoteness (in the FUTURE) of the predicated verb. That is to say, it does not matter whether the event will take place in the near future (as in (185) a.) or the distant future (as in (185) b. and (185) c.); any of the six constructions are valid for that purpose.

(T12.8.1.el)

- (185) a. **uvetřari** _____ **řrubk-u**
 in_evening [LIGHT UP] heating-ACC.SG

Accepted

majusj topitř [DE-OBLIGATIVE]
 topitřimu [SYNTHETIC]
 budu topitř [COPULAR]
 zatoplju [PERFECTIVE]

Rejected

*xotřu topitř [DE-VOLITIVE]
 *stanu topitř [DE-VENITIVE]
 #puřřfusj topitř [PSEUDO-DE-ANDATIVE]

‘This evening I am going to light up the stove/heating.’

- b. **za mřsjats** _____ **poli**
 in month.ACC.SG? [TILL] field-ACC.SG

Accepted

majusj řorati [DE-OBLIGATIVE]
 budu řorati [COPULAR]
 řoratimu [SYNTHETIC]
 vol. T10 > > zřoratimu [SYNTHETIC]

Rejected

*stanu řorati [DE-VENITIVE]
 *puřřfusj řorati [PSEUDO-DE-ANDATIVE]

‘In a month’s time I will till the field.’

c. **tʃrɪz** **ɦod** m-ij sin vid-je na
 in year.ACC.SG? 1SG.POSS-NOM.SG son.NOM.SG go.PRF-3SG in
 pens-ju i toɦɪ vɪn _____ v si'l-o
 pension-ACC.SG and then 3SG.NOM [MOVE] in village-ACC.SG

Accepted

maetsa pɪrɪjɪxatɪ [DE-OBLIGATIVE]

pɪrɪjɪde [PERFECTIVE FUT.]

Rejected

*pɪrɪjɪxatɪme [SYNTHETIC]

‘In a year’s time my son will be retiring and he will then move to the village.’

7.3.2.8. Restrictions as a complement of negated cognitive verbs

Based on Dahl’s (2000a) questionnaire, FUTURE TENSE forms can present restrictions when they appear as a complement of a negated cognitive verb (e.g. ‘know’), but there is no reason to believe that this applies to West Polesian. Nonetheless, it is true that the DE-OBLIGATIVE (i.e. *majusj(a)*) form was far more often rejected and replaced by the DE-VOLITIVE (i.e. *xotʃu*). The DE-VENTIVE future was constantly rejected, as in most of the grammaticalisation tests. As with the DE-OBLIGATIVE form, the reason for the rejection is almost certainly more related to the semantics of the verbs with which it was elicited.

(186) (Tor1.57.1el.) [Context: My granddaughter is going to start university]

ni ponimaj-u²⁵⁰ kudi vona_____

NEG understand-1SG where 3SG.NOM.F (ENROL)

majitsa postupiti/postupati [DE-OBLIGATIVE]

xotje postupiti [DE-VOLITIVE]

bude postupati [COPULATIVE]

postupatime [SYNTHETIC]

postupitj [PERFECTIVE FUT.]

*stane postupati [DE-VENITIVE]

‘I don’t understand where /what she is going to enrol in.’

7.3.2.9. Phonetic erosion/partial loss of inflection

It is interesting to remark that the process of phonetic erosion is still noticeable in at least two of the constructions, which adds more evidence to the suggestion that they are highly grammaticalised.

7.3.2.9.1. The SYNTHETIC FUTURE TENSE

The first person singular suffix *-imu* is being replaced for a shorter form *-im*. I have documented both forms, but a native speaker from Bahdanaŭka (B21) made a remark about this. According to her, older people say “*ja ro'bitimu*”,²⁵¹ whereas younger people (born after 1960, circa) say “*ja ro'bitm*” (B21.1 10:09). This

²⁵⁰ N.B. she consciously rejected the form *rozumju*, which seems the ‘most Polesian’ lexeme (compare also [CSR] *ponimaju* with [BLM] *razumeju*). In my experience, as already noted in (§2.3.3.) whenever language assistants tried to sound ‘more educated’ they tried to speak Russian, which often resulted in a speech, in which West Polesian lexicon, phonology and grammar was full of Russianisms.

²⁵¹ “Older people say ‘ja ro'bitimu’, ‘ja robitimu’ and we say... ‘ja robitm’. But forms like *robitr-mu*, [well] we employ them more rarely.” (B21.1 10:09) [My translation].

alternation is also present in T2's speech, who was 71 years old (in 2017), and who could be the bridge between older and younger people's varieties.²⁵²

Moreover, the SYNTHETIC FUTURE is in itself a phonetically eroded periphrasis of infinitive + **matr* or **imatr*.²⁵³ The fact that the phonological form of that auxiliary is so eroded that speakers cannot recognise it, and even different linguists argue about its etymology, proves that the grammaticalisation is very advanced. However, I discuss the etymology of this form in more detail in (§7.4.1.).

7.3.2.9.2. The DE-VOLITIVE FUTURE TENSE

The verb/auxiliary *xotiti* can appear without the PERSON/NUMBER inflection suffixes (or with a shorter form of these) in informal speech. Whenever I asked language consultants about this, they all pointed out that it is wrong. However, some other non-native speakers, such as their children, were able to immediately recognise it as a 'Polesianism' when I used it in an informal context when we were speaking Russian.

Most of the instances registered are from overheard conversations (especially 1SG forms) (187). The [1SG] *xotfu*, can be realised as > *xo*, *xotf*; and the [2SG] *xotfej* > *xotf*.

²⁵² Mackevič (1959: 201) also pointed out variation in the phonological realisation of the suffixes for person across Southern Belarusian varieties with the synthetic future (e.g. [1sg]-*mu* /-*mu*).

²⁵³ For this reason, it has been described as an equivalent of the DE-OBLIGATIVE form (e.g. *maju robiti*), which, after reanalysis, turned it into an inflectional suffix (DABM, 1963 vol. II: 662-672).

(187) a. (T2. OV) [quoting a man not willing to take part in the research]

Oj, **ni xo,** **ni xo,** **ni xo!**

Oh, NEG want.[1SG] NEG want.[1SG] NEG want.[1SG]

‘Oh, I don’t want to, I don’t want to, I don’t want to.’

b. (Z.4.1.1 09:36)

v mafiazin-ax je vs-jo ʃto **xotʃ**

in supermarket-LOC.PL be.PRS all-NOM.SG.N REL.ACC.SG want.[2SG]

‘In the supermarkets, there’s all you want.’

The eroded form of 1SG is less common than the form of 2SG. I have not been able to document this with other PERSONS and, as I have mentioned, it was not possible to directly elicit it during the session, as all participants denied using this form.

Neither phonetic erosion nor overdifferentiation are *sine quibus non* conditions for a split to happen. Having said this, the originally DE-VOLITIVE form is apparently developing a different inflectional paradigm. That is to say, I only documented the phonetically eroded paradigm of *xotitʃ* (e.g. [2SG] *xotʃ*) used on its primary volitive meaning (‘want to’). This can highlight the morphological robustness that the DE-VOLITIVE form is acquiring, and the presence of a lexical split.

7.3.2.10. Preservation of former meanings

At least half of the constructions under study here retain some semantic content from their original meaning, i.e. they have not yet completed the full process of desemanticisation. For some, this may be the strongest criterion or test for grammaticalisation; i.e. a structure cannot be fully grammaticalised unless there has been a process of ‘semantic bleaching’. Consequently, the constructions under study

are far from being grammaticalised. Nevertheless, we are seeing that semantics is not the only relevant criterion or test for the ‘canonicity’ of a form’s grammaticalisation. For example, the PERFECTIVE FUTURE is, in a certain way, the most grammaticalised form of all the analysed forms, because it only involves the verbal base and looks semantically neutral. Nonetheless, the verbal base upon which the PERFECTIVE future is built is derivational. Thus, if we only knew the PRESENT 1SG of a verb or its IMPERFECTIVE INFINITIVE we could predict the form of all the future constructions here except for the PERFECTIVE FUTURE. Moreover, the correlation between the IMPERFECTIVE and PERFECTIVE verbal bases is often not regular; i.e. the PERFECTIVE often adds extra semantic nuances (besides ASPECT) which are not present in the IMPERFECTIVE base. Given that this construction relies in derivation, at least to a certain degree, it shows a less grammaticalised nature than the rest of the constructions under study.²⁵⁴

7.3.2.10.1. The DE-VENITIVE FUTURE TENSE

As I have shown in (§7.2.4.2.) there is at least a homophonous form of the verb **statī*, meaning ‘to become’. On the surface, the DE-VENITIVE has often a strong inchoative flavour and can be often translated as ‘to begin’ (remember (156) a-b).

7.3.2.10.2. DE-VOLITIVE FUTURE TENSE

Besides its obligative/prospective meaning, the verb *xotitī* is still widely used as a plain volitive verb. This is true for both the varieties where it does not have that obligative/prospective acceptance (East), as well as for those where it does (West).

²⁵⁴ I discuss the PERFECTIVE FUTURE in more detail in (§7.4.5.).

(188) a. (Z.4.1.2 02:22)

bo treba bu-l-o, xoti-l-o-sja xlib-a
as need be-PST-N.SG want-PST-N.SG-REFL bread-GEN.SG

‘Because it was necessary, we wanted bread.’

b. (T11.5 00:27)

zare nits ni xotŭ-utj robi-ti!
now nothing NEG want-3PL work-INF

‘[Young people] nowadays don’t want to do anything.’

7.3.2.10.3. The DE-OBLIGATIVE FUTURE

Even though *DABM* (1963 vol. I: 166) described the existence of the DE-OBLIGATIVE FUTURE construction like *maju robiti* [NON-REFLEXIVE] in South Western Belarus, I was not able to find a village where the construction is not used with a REFLEXIVE. Whenever I asked speakers (who use *majusj*) about the NON-REFLEXIVE form, they recognised it, but they said it is used in other villages, which they could not name. Thus, for those who use the REFLEXIVE form, the NON-REFLEXIVE has a deontic meaning, contrary to what would be expected from the descriptions in *DABM* (1963 vol. II : 662-672).

(189) a. (T6.3 03:39.pr)

ja zautra kraj maj-u jixa-ti u bolnits-u
1SG.NOM tomorrow EMP have-1SG go-INF to hospital-ACC.SG

‘Tomorrow I must (irremediably) go to the hospital.’

b. (T2.ov)

ma-ju pi-ti jak traba bu-l-o pi-ti
have-1SG drink-INF like necessary be-PST-N.SG drink-INF

‘I must take [them], because [I was told] that I need to take [my pills].’

c. (Tor1.1.1 04:51.pr)

fiosudarstv-o pomofi-l-o [...] fiet-im trom brat-am
government-NOM.SG help-PST-N.SG that-DAT.PL three-DAT.PL brother-DAT.PL
ustroj-ti-sj, po...stroj-ti-sj seb-ji fjerm-u, tʃɪ tam
build-INF-REFL build.PRF-INF-REFL REFL-DAT farm-ACC.SG or there
ʃ:jo-nebudj, ʃ:jo **ma-je** **postroi-ti**
something.ACC.SG REL.ACC.SG have-3SG build.PRF-INF

‘The government helped these three brothers to build a farm, or whatever they had to build.’

Moreover, the verb *maju* in itself means ‘to have’ in each of the varieties analysed. It is not as frequent as the alternative periphrasis: *v* + possessor [GEN] + [*je(stja)*] + object [NOM], as also happens in BLM and ULM (Mazzitelli 2011).

(190) (B20.12 01:19)

xto firoʃ-i **ma-je**, to nɪtʃ nɪ sp-itj.
REL.NOM.SG money-ACC.PL have-3SG so at_night NEG sleep-3SG

‘Whoever has money, they don’t sleep at night.’

7.3.2.11. Recursivity

One of the signs of ‘semantic bleaching’ (i.e. a classical approach to grammaticalisation) may be that, since the original auxiliary has been desemanticised, the main verb can be the same as the one which gave rise to the auxiliary. However, in this case, not all the constructions that have an auxiliary in their base can create the FUTURE TENSE of the original meaning of that auxiliary. Thus forms such as: **budu buti* (**stanu buti?*); **stanu statı?* (**budu*

statı); and **xotfu xotıı* are ungrammatical.²⁵⁵ The PERFECTIVE cannot be included in this category, because it is not morphologically based on an auxiliary verb. The only exceptions are the DE-OBLIGATIVE (if we disregard reflexivity) and the SYNTHETIC forms.

(191) a. (Tor1.59.2.el)

mi **maj-ımo-sja** pjatero dıt-aj **ma-tı**
 1PL.NOM have-1PL-REFL five(COLL) child-GEN.PL have-INF

‘We are planning on having five children.’

b. (Tor1.59.2.pr) [Context: Asked about a pregnant woman and the gender of her future child]

ma-tı... **nı** **vjadom-o** kofıo **ma-tı-me.**
 have-INF NEG known-N.SG Q.ACC have-INF-HAVE.3SG

‘It’s unknown what [lit. ‘whom’] she is going to have.’

I have also found interesting examples in Klimčuk’s Literary West Polesian. For example, I note that in (192) c. the verb *matı*, which (synchronically) means ‘to have; to obtain’ has a SYNTHETIC FUTURE form.

²⁵⁵ There is a problem with the infinitive of **statı* because it does not exist in Contemporary West Polesian, and thus it cannot be elicited.

Klimčuk's (2010) translation of the NT (See footnote 220)

(192) a. (Matt 19:21)

Isus skaza-v jomu; “Jak xotŝ-ŝ bu-ti odpovidn-im,
 Jesus.NOM say.PRF-PST.M.SG 3SG.DAT COMP want-2SG be-INF righteous-INS.SG
 id-I, prod-aj majk-u svoj-u i rozd-aj
 go-IMP.2SG sell-IMP.2SG wealth own-ACC.SG and scatter-IMP.2SG
 ubofi-im; i **bud-ŝ** **ma-ti** skar^lbi na neb-I.”
 poor-DAT.PL and be.FUT-2SG have-INF treasure.ACC.PL in heaven-LOC.SG

‘Jesus answered, “If you want to be perfect, go, sell your possessions and give to the poor, and you will have treasures in heaven”.’

b. (Matt 27:22)

“I vse, tŝofo vite v molitv-I popros-ite
 and all-ACC.SG.N REL-GEN.SG 2PL.NOM in prayer-LOC.SG ask.PRF-2PL
 z vir-Iju, t-oe **mati-mi-te.**”
 with faith-INS.SG that-ACC.SG.N have-INF-HAVE.2PL

‘And anything you will pray for with faith, you will have it.’

7.3.2.12. Predictions vs intentions

Following Dahl's (2000b) predictions, the fact that certain constructions are currently restrained to intentions indicates an earlier stage of grammaticalisation or extension of the semantics of the FUTURE TENSE construction:

“[...] markers that are originally restricted to intention-based FTR [future time reference] tend to develop into general future markers, which include prediction-based FTR as central cases but can [...] still be used for intention-based FTR” (Dahl 2000b: 310).

For this reason, even though this is a test closer to semantics than morphology, this can reveal important information about grammaticalisation. Having said this, the division between predictions vs intentions is not necessarily clear cut and I have documented most constructions being used with both purposes. Nevertheless, in general, the DE-OBLIGATIVE and the DE-VOLITIVE align better with intentionality, whereas the SYNTHETIC FUTURE corresponds with predictions.²⁵⁶

I had noticed that T8 and T14 were using the SYNTHETIC FUTURE to talk about the rain (as in (172)). I also tried eliciting the construction with the COPULAR form, as well as the DE-OBLIGATIVE from Tor1. She approved all of them, but with the DE-OBLIGATIVE, she reacted this way (193):

(193) (Tor1.59.3.el)

obiʃtʃa-l-I ʃtʃo doʃtʃ xoroʃ-I, **maj-it-sja** **bu-ti** doʃtʃ
 promise-PST-PL COMP rain.NOM.SG good-NOM.SG have-3SG-REFL be-INF rain.NOM.SG

‘They (had) promised that it will rain.’

I asked whether there is any difference between some people promising it and us being totally certain that it will rain. She said that in both cases it is still correct.

²⁵⁶ This last point is more of an impressionistic analysis based on a limited corpus. No Belarusian linguist has paid much attention to the analogous form in Southern Belarusian, but those who have (Jankoŭski 1989, Mackevič 1959, Mackevič et al. 1964) have been unable to find any semantic differences between the COPULAR and the INFLECTIONAL forms. The same can be said about the Ukrainian SYNTHETIC FUTURE (Mikhaïlyk 2003, Pugh & Press 1999, Rusanovskij et al. 1986, Shevelov 1963).

The PERFECTIVE FUTURE and the COPULAR construction are more universal, as in some varieties only those two and the SYNTHETIC FUTURE TENSE exist, yet they are better for describing predictions than intentions.

Predictions

(194) a. (B21.6 06:00)

bu-l-i prorotfestvovujstf-ie, to jomu [...] bu-l-o tak-eje
be-PST-PL foretelling-NOM.PL so 3SG.DAT.M be-PST-N.SG that-NOM.SG

slov-o ot Bofi-a, ftfo [...]vin [...] tak-oju smert-eju
word-NOM.SG from God-GEN.SG COMP 3SG.NOM.M that-INS.SG.F death-INS.SG

pomr-e. no mi 3 tofdi ni zna-l-i k tjomu fiet-o
die.PRF-3SG but 1PL.NOM PART then NEG know-PST-PL for Q.DAT that-NOM.SG.N

bu-l-o, fiet-o slov-o, ftfo [...] tak **polutš-et-sa.**
be-PST-N.SG that-NOM.SG.N word-NOM.SG COMP so happen.PRF-3SG-REFL

‘There were some people that prophesied; he received such word from God that he was going to have such a death. But then we didn’t know why [they had said] that word, that that was going to happen.’

b. (TL6.2 01:15)

ftfo **prisn-it-sja** to je i **ispoln-it-sja**
REL.NOM.SG dream_of.PRF-3SG-REFL that_ same-NOM.SG so fulfil.PRF-3SG-REFL

[A ‘prophecy’] ‘Whatever you’ll dream of, it will happen.’

c. (B7.5 03:20)

bo ja doufio ni bud-u, ja **vmr-u**, [...], bo ja boln-i
as 1SG.NOM long NEG be.FUT-1SG 1SG.NOM die.PRF-1SG as 1SG.NOM sick-M.SG

‘I will not be around much longer, I will die, because I’m sick.’

d. (B6.20 06:12)

mi [...] znaj-em "ve'silje", a jak kto, to **ni zna-ti-me**
1PL.NOM know-1PL *vesilje* but if Q.NOM so NEG know-INF-HAVE.3SG
'We know [what] "ve'silje" (wedding) [means], but someone [else]
won't know it.'

Intentions

(195) a. (T11.21 00:45.pr)

haz **ma-l-i-sja privis-ti**, jak xtos ja zabra-v.
gaz.ACC.SG have-PST-PL-REFL bring.PRF-INF if so 1SG.NOM take.PRF-PST.M.SG

'They were supposed to bring the gas, so I would have gone for it.'

b. (TL1.2 01:57)

a skaz-i [...] mo tak "po pravd-e, **ni bud-em** boljše
so say.IMP-2SG maybe like for truth-DAT.SG NEG be.FUT-1PL more
ruha-ti-sja". No i vs-i skaza-l-i po pravd-e, "i
curse-INF-REFL so and all-NOM.PL say.PRF-PST-PL for truth-DAT.SG and
bilš vze ni ruha-ti-mo-sj", i pomiri-l-i-sj
more already NEG curse-INF-HAVE.1PL-REFL and reconcile.PRF-PST-PL-REFL
'And say something like "we promise we will not curse again". So then
they all promised "we will never curse again", and they reconciled.'

c. (Z10.4 00:23)

ja vam **roskaž-u** zare t-eje stixotvorenj-e,
1SG 2PL.DAT tell.PRF-1SG now that-ACC.SG.N poem-ACC.SG

ʃ:tʃjo ja sklada-l-a.
REL.ACC.SG 1SG.NOM compose-PST-F.SG

'Now I will tell you that poem that I wrote.'

In sum, the SYNTHETIC, the PERFECTIVE and the COPULAR future are more grammaticalised than the rest, in this respect, because they can be used to express predictions.

7.3.3. Summary

According to Bybee (1985: 158-159) the etymology of future forms (or affixes) is usually traceable, whereas this is often not the case with PAST TENSE inflections cross-linguistically. She points out that, according to different surveys, the most common sources for the development of FUTURE TENSE forms are “constructions expressing *obligation or necessity, desire, and movement or intention*” (Bybee 1985: 158). Indeed, the fact that most FUTURE TENSE constructions here retain modal nuances is evidence that the grammaticalisation of the FUTURE TENSE constructions is recent and still traceable. Therefore, we see this as a natural or expected process, which brings more solid evidence for treating these FUTURE TENSE constructions as grammaticalised.

It must be said that the COPULAR, the PERFECTIVE and the SYNTHETIC future tense were documented in all the varieties surveyed. The DE-OBLIGATIVE and the DE-VOLITIVE were only documented in the West. And the DE-VENITIVE was primarily documented in the West, although there are some hits on the corpus of the East.

I have summarised the results of the grammaticalisation tests in Table 39. The more + a particular FUTURE TENSE construction obtains from the diagnostic tests, the closer it is to the ‘canonical’ grammaticalised FUTURE TENSE construction.

Table 39 Summary of the main grammaticalisation tests applied
(+/- = relevance)

	SYNTHETIC FUTURE	DE-OBLIGATIVE FUTURE	DE-VOLITIVE FUTURE	DE-VENITIVE FUTURE	PERFECTIVE FUTURE	COPULAR FUTURE
1. Segmentable	N/A	+	+	+	N/A	+
2. NO PERSON restrictions	+	+	+	+	+	+
3. NO ASPECT restrictions	-	?	+	?	-	-
4. NO TRANSITIVITY restrictions	+	+	+	?	+	+
5. NO MOOD restrictions	+	+	?	?	-	-
6. Available for CONDITIONAL protasis	+	+	?	?	+	+
7. NO EVENT-TIME restrictions	+	-	-	-	+	+
8. NO restr. as c. of negated cognitive verbs	+	+	+	?	+	+
9. Phonetic erosion	+	-	+	-	-	-
10. NO additional meanings	+	-	-	-	+	+
11. Recursivity	+	+	-	-	-	N/A
12. Available for predictions	+	-	-	+	+	+

7.4. West Polesian future-tense constructions from a historical-typological perspective

In this section, I run through the (potential) six main constructions identified at the beginning of this chapter, in line with Dahl's (2000b) survey. This discussion is almost entirely based on Dahl's survey and only serves the purpose of contextualising the development of futures cross-linguistically. I start by analysing the relations each construction may have to its closest Slavonic relatives and the chronology of such forms in Slavonic, progressively moving towards the 'bigger European picture'. On the one hand, this will provide an empirical base for the final verdict on the nature of the constructions discussed in this chapter. I show that these constructions have parallels in other European languages (which are not necessarily related); and so that those semantic bases that were more doubtful, also meet the typological expectations. On the other hand, this will also emphasise that it is not surprising to see several different FUTURE TENSE constructions in one language (see Bybee's (1985) comments in (§7.1.2.)), as well as very different strategies in otherwise genetically closely related languages. Andrew Spencer (p.c.) illustrates this with Germanic languages. Notwithstanding their genetic affiliation, they have developed very different types of FUTURE constructions; e.g. Swedish *skulle* [pseudo-DE-OBLIGATIVE] and *ville* [DE-VOLITIVE]; English *shall* [DE-OBLIGATIVE] and *will* [DE-VOLITIVE]; or German *werden* [DE-VENTITIVE].

7.4.1. The SYNTHETIC FUTURE

There is no reason to reject the idea that the SYNTHETIC FUTURE in West Polesian and the SYNTHETIC FUTURE that is widespread in Ukrainian and some Southwestern Belarusian dialects (particularly those close to West Polesian) have a common origin (DABM 1963, Danylenko 2011, Mackevič 1959, Mackevič et al. 1964).²⁵⁷ In this subsection I focus on descriptions and analyses of this form in Southwestern Belarusian and Ukrainian and I point out whenever West Polesian differs from the former two.²⁵⁸

According to Pugh & Press (1999: 229), in Contemporary Ukrainian “[t]here is no functional or semantic difference” between the COPULAR and the SYNTHETIC FUTURE constructions, besides the fact that “the synthetic tends to be used less frequently than the analytic [COPULAR FUTURE], especially in West Ukraine”.

Based on the results from the analyses in the previous sections, the same can be said of West Polesian. There are apparently no differences between the COPULAR and the SYNTHETIC FUTURE TENSE, other than that the SYNTHETIC FUTURE form is not as frequent (see (§7.2.1.)). This may be an indication that the SYNTHETIC form is slowly receding, in favour of the COPULAR form, which the norm in BLM and CSR. According to Jankoŭski (1989), this form is rarely found in Standard Belarusian (BLM) since the

²⁵⁷ “In Southwestern Belarusian dialects besides the form built with the personal forms of the verb *byc* ‘to be’ + infinitive, there exists a SYNTHETIC form for the future tense, which preserves vestiges of the auxiliary verb *imu* which directly attaches to any infinitive and becomes a suffix (*rabic’mu*, *rabic’meš*, *rabic’me*, etc.)” (Mackevič 1959: 201) [My translation].

²⁵⁸ Irrespective of my position about Danylenko’s views, his description of this form (Danylenko 2011: 173) also includes the area where West Polesian is spoken, although no specific data for West Polesian are presented.

reform of 1933 which qualified the SYNTHETIC form as “dialectal”.²⁵⁹ Conversely, in (§7.2.1.) we have seen that West Polesian disprefers the SYNTHETIC form with reflexive (or polysyllabic) verbs, but that the REFLEXIVE appears after the inflection for TENSE (FUTURE) and PERSON. Nevertheless, according to Mackevič (1959: 206-207) there can be ‘infractions’ to the rules of the formation of the FUTURE TENSE in Southwestern Belarusian varieties; e.g. “*budu robĩtmu* [‘I will know’], [...] *xaj ja budu znatsjmu* [‘Shall I know!] (Slonim, 499) [...] *jtfo mnje tsjepjer robĩtsmu, kolĩ njema nofĩ* [‘what will I do now that I’m missing a leg’]”.²⁶⁰ The result is a redundancy in the marking of PERSON and TENSE.

Dahl (2000b: 319) says that the SYNTHETIC FUTURE in Ukrainian is based on auxiliary ‘have’, which leads him to suggest a possible link with the Balkan DE-OBLIGATIVE construction-family. This also encourages him to find a potential typological (not genetic) link with the DE-OBLIGATIVE FUTURE in Romance languages. Nevertheless, he admits that this particular grammaticalisation (or cliticisation) is an innovation within the Slavonic family. In fact, he claims that the Ukrainian SYNTHETIC FUTURE-tense has “[...] gone further in grammaticalization and differs aspectually from others” (Dahl 2000: 323).

²⁵⁹ I should add that the 1933 reform (which gave origin to the standard usually referred as *Narkamoŭka*) allegedly tried to bring Belarusian closer to Standard Russian. However, there are big controversies over the two main standards of Belarusian, which I do not intend to discuss here. See more details in Mazzitelli (2011).

²⁶⁰ Note that the author did not use the standard spelling for BLM, rather he was using the available orthography of Eastern Slavonic for his transcriptions. Thus, for this reason, instead of transliterating with the correspondences established for BLM, I have rendered the Southwestern Belarusian and West Polesian forms using the writing conventions I have settled for West Polesian through this entire work.

Danylenko (2011) argues against Dahl's interpretation of the FUTURE suffix as an apocopated form of *iměti* 'to have', and thus, denies any affiliation (genetic or typological) with the development of the DE-OBLIGATIVE FUTURE TENSE in Romance or any other 'have'-type (DE-OBLIGATIVE) future construction families.²⁶¹ Danylenko claims that:

“the Ukrainian s[ynthetic] f[uture] is a continuation of the de-inceptive p[eriphrastic] c[onstruction] with a weak grammaticalization of the auxiliary *jati* (< **jěti*) 'to take' historically undergoing grammation along the clitic continuum as [...] postulated elsewhere for the Ukrainian-speaking territories” (Danylenko 2011: 161).

According to Danylenko (2011: 161, 169) the overlapping of these alleged verbal forms, *iměti* 'to have' and *imati* 'to take' (respectively from **jīmō* and **jīmati*) took place already as early as the sixteenth century. This would have been partly motivated by the loss of the initial *j-* (which Danylenko (2011: 171-172) considers an independent development), which contributed to the desemanticisation of the verb **jati* 'to take'. So he claims that the auxiliary verb *jati* took its current form (i.e. cliticised/suffixed to the infinitive) in Ukrainian by the mid-nineteenth century (Danylenko 2011: 170).

In spite of his claims, Danylenko recognises the existence of a DE-OBLIGATIVE construction in earlier and middle stages of Ukrainian and Belarusian: “[d]espite its

²⁶¹ In defence of Dahl, it must be said, that in his works he acknowledges that the auxiliary 'have' (or modality) is a very common source for future tense constructions (Bybee & Dahl 1989, Dahl 2000b), and so he does not claim any genetic affiliation between these forms. Thus, some of Danylenko's (2011) arguments against Dahl's (2000b) analysis seem a bit of a straw man.

frequency, Middle Ukrainian (and Middle Belarusian) de-obligative constructions with *iměti* did not show ‘much futurity,’ retaining its underlying modal semantics” (Danylenko 2011: 166).²⁶²

Although I will not discuss Danylenko’s position in detail, I must mention some objections to his analysis. First, Danylenko (2011: 178) admits that the alleged verb “*jati* (< LCS [Late Common Slavonic] **jęti*) ‘to take’ [is] not found in West and South Slavic”, which makes his argument already suspicious. Moreover, this verb (if it ever existed) would have not survived (at least fully) in any Contemporary Slavonic variety.

Second, his historical evidence for proving the existence of an original inceptive verb *jati* is very arguable. The only possible two examples of a verbal form of the alleged *jati* come from a fifteenth-century Middle Ukrainian text, with two tokens of *imutъ* [3PL]. Even Danylenko (2011: 169) admits that that form could have either belonged to the paradigm of *iměti* ‘to have’ or *jati (imati)* ‘to take, to seize’. Thus, if there are not surviving attestations of *jati (imati)* (as ‘to take, to seize’) and it was allegedly being grammaticalised, is there reliable way of telling what it ‘really’ used to mean?

Third, I find that his morphological argument to justify the semantics of the verbal form (“inceptive”) is not logical:

²⁶² Now the Historical Dictionary of Belarusian (hereafter, HSBM vol. 14) documents the meaning of ‘to take; to seize’ as two of the meanings of *imati*; yet it also describes certain uses of it that suggest that it could be also the verb ‘to have’, as well as an auxiliary for a periphrastic future with a modal/deontic meaning (as early as 1438).

“There are no solid grounds for identifying the S[ynthetic] F[uture] as synthetic since the synthetic principle applies to the auxiliary clitic only, which reveals its primary inceptive meaning” (Danylenko 2011: 176).

Now, we have seen from the data in West Polesian (and neighbouring Southern Belarusian dialects, (Roncero 2015)) that there is a DE-OBLIGATIVE FUTURE TENSE construction form still alive, and according to Danylenko’s (2011: 172) comments, it would have also been preserved in Rusyn (for him Carpathian dialects).²⁶³ Therefore, is it possible to argue that the SYNTHETIC form is derived from a different auxiliary (allegedly meaning ‘to seize’), if in some varieties the auxiliary *matı(sj(a))* appears before the verb, making the structure very transparent?

On the one hand, in favour of Danylenko’s (2011) hypothesis, we would have to explain why *majusj(a) robıı* and *robıımu* are expressed differently if they both derive from the same auxiliary. According to Bybee (1985):

“[c]ases of reordering of morphemes are not very common, so it will be often the case that morpheme order reflects an earlier order of words, but it is still important to recognize that morphology is not immovable fossilized syntax. Speakers will sometimes rework parts of their morphology” (Bybee 1985: 41).

On the other hand, there is evidence of such a mobility in Slavonic. The (DE-VOLITIVE) markers BCMS uses for its FUTURE TENSE can appear either as suffixes or as enclitics,

²⁶³ “In archaic Central Transcarpathian dialects, this future is attested alongside with the PF *budu* + INF, e.g., *mu spıvaty* next to *budu spıvaty* ‘I shall sing’. Yet, what is remarkable about the western Ukrainian PF is that the auxiliary clitic ‘to take’ occur today in clause second or verb-adjacent position” (Danylenko 2011: 172). Nevertheless in Pugh’s (2009) Rusyn grammar there is no mention of this form. He only points out that “the alternate Ukrainian *synthetic* imperfective future, consisting of the fusion of the infinitive + *m* + personal endings [...] is *not* found in Rusyn” (2009: 139).

with different rules regarding their order (Alexander 2006). That is to say, given the peculiar nature of the FUTURE TENSE marking (particularly in Slavonic), it would not be surprising to discover that at some point in history both word orders were possible (i.e. the auxiliary before and after the verb). In fact, Jankouški (1989: 231) claims that in older Belarusian (or Eastern Slavonic) texts the auxiliary *imu* is found before and after the infinitive (although he does not provide any examples or sources). We can propose then, that with time, the word order which had the auxiliary after the verb would have been phonologically reduced and would have started to merge with the infinitive, due to the tendency for suffixation (rather than prefixation) of Slavonic. This way, the pre-verbal auxiliary form may have survived almost intact (although developed a new semantic nuance: intention), whereas the post-verbal form is shrinking (given that it still causes problems to speakers when they want to use it with long and/or reflexive verbal bases).

Finally, in addition to this, based on purely typological expectations (Bybee 1985, Bybee & Dahl 1989, Ultan 1978), a verb meaning ‘have’ is a far more likely semantic base for a FUTURE TENSE construction than a verb meaning ‘to seize’.²⁶⁴

The truth is we cannot know whether there were two verbal forms (i.e. one meaning ‘to have’; and another one meaning ‘to take/seize’) or not. Danylenko may be one of the latest scholars (and one of the most vocal) in proposing the de-inceptive origin of

²⁶⁴ The ‘DE-INCEPTIVE’ type of FUTURE TENSE constructions (i.e. those based on ‘seize; grab’) are rare cross-linguistically, yet Dahl (2000b: 324) very briefly mentions three examples from Hungarian, Romani and Turkish.

the SYNTHETIC FUTURE form, but according to Whaley (2000) there has been an ongoing debate around this:²⁶⁵

“At a very early period, the verbs *imatb* ‘have (to)’ and *imu* ‘take’ were confused in texts (Kuznecov 1959: 236; Kiparsky 1967: 234), most likely due to the fact that they are formed from the same stem (cf. Townsend and Janda 1996: 215-16). It is unclear which verb developed into the SYNTHETIC FUTURE found in Belarusian [sic] and Ukrainian” (Whaley 2000: 153) [See references there].

In sum, the SYNTHETIC FUTURE TENSE constitutes a construction-family or sub-family (or ‘gram-family’ using Dahl’s terminology) in South Eastern Slavonic.²⁶⁶ There are two hypotheses about the original etymology of the suffix attached to the infinitive. One of them (defended by Danylenko) says that it is a verb meaning ‘to take, to seize’. The other, which I find more convincing, argues that it is the same verb ‘to have’ (reduced into a suffix), found as a modal auxiliary as well as a DE-OBLIGATIVE form in old and modern Slavonic, including West Polesian; and which aligns much better with typological expectations. Regardless of the approach we take (i.e. whether we believe the SYNTHETIC form is derived from the verb ‘to have’ or ‘to take’) its areal or typological links with other forms (e.g. Bulgarian/Macedonian DE-OBLIGATIVES) are weaker, in comparison with the rest of the FUTURE TENSE constructions under discussion. Thus, it seems reasonable to propose that this form is a South Eastern Slavonic innovation. Some Southwestern Belarusian dialects have gone a bit further

²⁶⁵ For example, Kozhanov (2016) also believes that the verb **jati* originally meant ‘to take’, and that it became the source of the periphrasis in “Old Russian”. His main argument is that Russian Romani has developed a FUTURE-TENSE AUXILIARY based on the verb ‘to take’, which would be an old calque from Russian.

²⁶⁶ That is to say, covering South-western Belarusian, Polesian and Western Ukrainian, mostly.

along the grammaticalisation pathway, allowing double marking (Mackevič 1959).²⁶⁷ Conversely, West Polesian is marginally innovating, with regards to the rest of the group, in that it has started allowing PERFECTIVE verbal bases with it (as I showed in examples).



Figure 11 Distribution of SYNTHETIC FUTURE in Europe. Sources (Dahl 2000b, Danylenko 2011, Mackevič 1959).

²⁶⁷ Mackevič (1959) also provides examples of ‘marking mistakes’ (i.e. instability of the form) when using the inflectional future in South Western Belarusian varieties.

7.4.2. The DE-OBLIGATIVE (or ‘have’) FUTURE

In the villages where I have worked I have only documented the reflexive form of the auxiliary (*matysj(a)* + INF) ‘to have’. Nevertheless, when I asked my language assistants, they all recognised the non-reflexive form (*matr* + INF) as grammatical. They pointed out that it is used in other villages, yet nobody managed to specify a village/variety where they use it. In any case, even though I primarily deal with the reflexive form here, the semantic base of the auxiliary is the verb ‘to have’ and so I include it in the group of DE-OBLIGATIVE FUTURE constructions.

Before continuing further, in the previous sub-section (§7.4.1.) I mentioned that Dahl (2000b: 323) suggested a possible link between the (Ukrainian) SYNTHETIC FUTURE and the Balkan DE-OBLIGATIVE form.²⁶⁸ Since West Polesian has both forms, I deal with this form separately from the SYNTHETIC form (see (§7.4.1.)) admitting that there may be some links between the two, but their grammaticalisation paths have been very different.

7.4.2.1. Eastern Slavonic links

Not all the contemporary Slavonic languages have preserved a verb for ‘to have’ (particularly in the East), but there are obligative constructions in many, which use the verb ‘to have’ ((i)*miec* or any of its variations); e.g. Polish *Mam coś robić* ‘I have to do something’. Neither have all the contemporary Slavonic languages

²⁶⁸ Even though Dahl (2000b: 323) calls this gram-family the Balkan ‘have’ future, for the sake of consistency with the rest of the terminology, and in order to ease the comparison with analogous constructions from other areas of Europe, I call this future gram ‘DE-OBLIGATIVE’ (see footnote 224).

grammaticalised (or preserved) the verb ‘to have’ as a FUTURE TENSE auxiliary, but there are a few.

According to Danylenko (2011: 172) and Whaley (2000:58) Rusyn or “Transcarpathian dialects” have a DE-OBLIGATIVE FUTURE form, which is realised analytically, the auxiliary appearing before the infinitive. Nevertheless, Pugh (2009) does not make any reference to this form in his grammar of Rusyn. Whaley (2000: 61) briefly mentions that apparently there are DE-OBLIGATIVE FUTURES in North Russian dialects (more specifically she mentions the varieties spoken in the Velikij Ustjug, Grjazovec, Kadninkov and Čerepovec regions), although she does not provide any examples.

Belarusian (BLM) has the closest analogue to the West Polesian DE-OBLIGATIVE. In both Standard Belarusian (particularly in Narkamaŭka) and West Polesian the verb (BLM) *mec’*/ (WP) *mati* ‘to have’ is almost extinct. Mazzitelli (2011) ran a corpus survey on both *mec’* and *mecca* (the reflexive form) in Contemporary Standard Belarusian texts.²⁶⁹ Given that the verbs are rare *per se*, not surprisingly, the results she obtained were that both forms are very infrequent, particularly the reflexive one: 0.05 per thousand words for *mec’*; and 0.01 per thousand words for *mecca* (2011: 180-181).²⁷⁰ According to Mackevič (1959: 200) in Standard Belarusian, both *mec’* and *mecca* + INFINITIVE can be employed in the PAST and the PRESENT TENSES.

²⁶⁹ Lomtev (1956: 181) says that, although very rarely, constructions with *mec’* can be found in Contemporary BLM works and that they used to be the norm (and where widely used) in older stages of Belarusian. He claims that nowadays such constructions should be considered “dialectal”.

²⁷⁰ It must be also said that the reflexive form *mecca* is equally frequent or infrequent in both standards of Belarusian (*Taraškevica* and *Narkamaŭka*), whereas the contrast in the use of *mec’* is more notorious between the standards (Mazzitelli 2011: 181).

Now, according to Mackevič (1959: 200) and Mazzitelli (2011: 182) in Belarusian, both *mec'* and *mecca* combine with PERFECTIVE and IMPERFECTIVE infinitive forms and they do not impose semantic restrictions on the subject; i.e. they allow either ANIMATES or INANIMATES to be subjects of the verb. Thus, the West Polesian and BLM DE-OBLIGATIVE forms share a great deal in common, and it is reasonable to infer a common origin.²⁷¹

When it comes to the uses and meanings of both *mec'* and *mecca* + INFINITIVE, Mackevič (1959: 200) says that in Standard Belarusian they can have different modal meanings:

“The modal meanings vary according to the form of the verb used: *mec'* or the reflexive *mecca*. The first one expresses a stronger meaning of obligation, need. [...] Constructions with the reflexive form, *mecca*, are characterised by a modal meaning of intention, disposition to undertake the action. [...]

The (modal) semantic nuances of these constructions are close to modal constructions of the type *xacec'* [‘want’], *žadac'* [‘wish’], *dumac'* [‘think’], *music'* [‘must’]” (Mackevič 1959: 200) [My translation].

On the one hand, the BLM form *mec'* lies between temporality and modality. Among the temporal notions it can express (mostly FUTURE), there are scheduled future (the most common), fatalistic future (especially when used in the PAST TENSE) and intentionality (Mazzitelli 2011: 182, 190).

²⁷¹ I lack enough data to assert whether West Polesian imposes any animacy restriction on its subjects with DE-OBLIGATIVES. Nevertheless, I have observed ANIMACY restrictions on the object of the DE-OBLIGATIVES (see (§7.2.2.2.), particularly (150)).

BLM [*Naša Niva* (2007), In Mazzitelli (2011: 182)]

- (196) Imprez-a **me-l-a** **adby-c-ca** 13 ljut-aha,
 party(f)-NOM.SG have-PST-F.SG happen.PRF-INF-REFL 13 February-GEN.SG
 ale administracyj-a admovi-l-a ŭ jaho pravjadzen´n-i.
 but administration-NOM.SG ban.PRF-PST-F.SG in 3SG.ACC.N celebration-GEN.SG
 ‘The party was due to take place on the 13th of February, but the
 administration banned its celebration.’

Very rarely, BLM *mec´* can be used to express immediate future, but retaining a flavour of obligation, which for Lomtev (1956: 181) and Mazzitelli (2011: 184) disqualifies it as a fully grammaticalised (i.e. in their understanding, semantically neutral) FUTURE auxiliary. Allegedly, the used of BLM *mec´* can be also linked to evidentiality. That is to say, *mec´* is more frequent with reported speech, as a type of reportative (Mazzitelli 2011: 185).

On the other hand, Mazzitelli (2011: 185) says that the primary function of Contemporary BLM *mecca* (i.e. the reflexive form) is “to express the notion of scheduled future in the past, usually with counterfactual interpretation [...], and intention”.²⁷²

BLM [See references in Mazzitelli (2011: 181)]

- (197) Ja **me-l-a-sja** **sta-c´** povar-am
 1SG.NOM have-PST-F.SG-REFL become-REFL cook-INS.SG
 ‘I should have become a cook.’

²⁷² By counterfactual she means an event which was about to happen, but which in the end did not.

Notwithstanding the semantic differences between the two forms, Mazzitelli's (2011: 180-181) survey showed that *mec'* is primarily used in the PRESENT TENSE, whereas *mecca* (i.e. the reflexive form) almost exclusively in the PAST TENSE. That is to say, the main condition for one form or another seems to be TENSE. In contrast, as has already been shown, West Polesian *matisja* + INF appears more often in the PAST TENSE (either as a counterfactual event or an intention). Yet it is still quite frequently employed in the PRESENT/FUTURE TENSE (i.e. as a proper future tense auxiliary), with an added meaning of **intentionality**. That is to say, the West Polesian reflexive auxiliary is not TENSE-bound, and displays more traits of a canonical or grammaticalised FUTURE TENSE construction.

7.4.2.2. Chronology in Eastern Slavonic

According to Whaley (2000: 54) “[...] the fourteenth-sixteenth centuries show widespread use of *maju* plus the infinitive, a ‘have/take’-type future similar to that found in Old Church Slavonic as well as Rus’ian Church Slavonic”.²⁷³ By contrast, the reality I have found in *The Belarusian Historical Dictionary* (henceforth, HSBM)²⁷⁴ is quite different. I have searched for all the possible variants of the verb (*mati*; *imati*, *meti*) with their respective reflexive counterparts in HSBM. For the form *mati* as ‘to own; to have’, there are only a few historical attestations of the verb and not much

²⁷³ Conversely, the COPULAR FUTURE TENSE was marginal at that time (*ibid*).

²⁷⁴ For its abbreviation in Belarusian (*Historyčny Sloŭnik Belaruskaj Movy*).

information is given. However, I have identified the form in another text (but in an unrelated entry) where it was used with a modal value.²⁷⁵

When it comes to *imati* and *imatisja*; *HSBM* very briefly says that both have been used as auxiliaries for the analytic FUTURE TENSE. The earliest instance of *imati* that it documents is from 1438, where it says it has a modal/deontic nuance. And for *imatisja*, the earliest token also comes from a fifteenth-century text (although no comments are made about any additional nuances).²⁷⁶

Finally, the pair *meti* and *metisja* are the most prominent in the dictionary (*HSBM* vol. 18), and yet there is hardly any information available. Out of the many meanings for *meti*, it says that it was possible to use it as an auxiliary for FUTURE TENSE denoting obligation or possibility. The earliest potential example of a DE-OBLIGATIVE FUTURE that it provides is from a text dated from 1533, but whereas the deontic reading is very clear, the temporal value is highly disputable. Mazzitelli (2011: 187) goes even further saying that *HSMB* does not provide any legitimate example of *meti* as a FUTURE TENSE auxiliary. When it comes to *metisja* the available material is even more scant. Only the 7th entry describes very concisely that it has been used as a FUTURE TENSE auxiliary but denoting obligation. It only cites an example from 1579 (198):

²⁷⁵ In (*HSBM* vol. 32):

обоѡ	сторона	маѡтъ	перед	нами	оѡчиѡисто	стати
oboja	storona	ma-etъ	pered	nami	očivist-o	sta-ti
both.F	side(F).NOM.SG	have-3SG	before	1PL.INS	testimony-ACC.SG	stand-INF

‘Both sides must testify before us.’

²⁷⁶ According to this, first attested in the *Vislicki Statut*, whose manuscript has been dated to be from the XV century, although no more details are given.

(198) **Old Belarusian** (AVK, 406 (XVI century) In: HSBM vol. 18)

by-l-a u mene t-aja prijelatka-a ... o t-om
be-PST-F.SG in 1SG.GEN that-NOM.SG.F girlfriend-NOM.SG.F about that-LOC.SG

ne veda-l, esli by ee **me-l-i-se** **svata-c'**
no know-PST.M COMP PART 3SG.ACC.F have-PST-PL-REFL propose-INF

ljud-i abo ne
person-NOM.PL or no

'I used to have a girlfriend... I didn't know whether they were going to offer her in marriage or not.'²⁷⁷

In the light of the data from HSBM, we can deduce that the form *meti* and particularly *metisja* were rare in Old Belarusian, as Mazzitelli (2011: 190) did.

I must admit that the path of development of the form *mec'/matr* (and variants) in Eastern Slavonic is more or less documented (and perhaps not very difficult to trace). However, the origin of the reflexive form *mecca/matisja* (and variants) is still an unresolved question and, so far, little attention has been paid to it.²⁷⁸ Mazzitelli (2011) points out that the reflexive form was already present in Old Belarusian, and it already differed from its neighbours:

“[...] it should be taken into account that Old Belarusian syntax has been influenced extensively by Polish, and, indeed, the use of *mec'* + infinitive in Old Belarusian is often the same as the use of *mieć* + infinitive in Old Polish [...] On the contrary, no Polish influence can be evoked for *mecca*, as Polish *mieć się* is never used governing an infinitive” (Mazzitelli 2011: 186).

²⁷⁷ Could be also understood as 'to propose to her'.

²⁷⁸ Part of the reason could be that in BLM *mecca* is rather a marginal form and dialectal connotation (in contrast with *mec'*) (based on Mazzitelli's (2011) description of BLM *mecca*).

According to Mazzitelli's corpus study (2011: 192-193, 199), Old Belarusian *mecca* was already grammaticalised as an auxiliary; and it was available for the FUTURE and the PAST TENSES (as in West Polesian), the FUTURE being the least common. Yet, already by the nineteenth century it was almost restricted to the PAST TENSE (*mec'* being used for the future, instead), expressing intention or mostly a counterfactual meaning (Mazzitelli 2011: 192, 193, 199). The reflexive PAST TENSE form was preserved into the twentieth century in BLM. The use of *mecca* did not decrease as significantly as the BLM *mec'* in Soviet Belarus, but it remained very much marginal.

Mazzitelli (2011) presents various theories about the grammaticalisation path of the reflexive DE-OBLIGATIVE auxiliary. She recognizes that

“[i]n Old Church Slavonic and in most old Slavic languages a form of reflexive ‘have’ is testified, but it has mostly the meaning of ‘existing,’ ‘behaving’ and ‘feeling (good, bad)’” (Mazzitelli 2011: 198).

Therefore, Mazzitelli thinks this is an unlikely semantic base to develop a FUTURE TENSE auxiliary (in the light of Bybee's (1985) study). According to her, one possible explanation is that the auxiliary may have emerged from one of the meanings the reflexive *mecca* had in Old Belarusian ‘strive for, going to’ (without an infinitive) (Mazzitelli 2011: 198).

However, the theory that Mazzitelli (2011: 199-200) favours most is that Old Belarusian *meti* and *metisja* emerged almost in parallel. With time *meti* would have retained some modal value and added a value of ‘external obligation’, whereas *metisja* would have developed into a more neutral auxiliary to replace the gaps of *meti*,

particularly in the PAST TENSE. In addition to this, Mazzitelli finds this last explanation particularly helpful, because it helps explain why the reflexive forms are overwhelmingly more common than the non-reflexive ones in all periods of Belarusian when used with FIRST and SECOND PERSONS.

In the light of this, we can say that West Polesian diverged from the rest of the Eastern Slavonic varieties already by the sixteenth century by preserving the reflexive form as a FUTURE TENSE auxiliary. So, in this respect, West Polesian is more conservative than the rest of the Eastern Slavonic sub-family.

7.4.2.3. DE-OBLIGATIVES as a Balkan phenomenon

In order to find other relatives of the DE-OBLIGATIVE FUTURE, we have to move onto more distant members of the Slavonic family. Dahl (2000b: 323) described the DE-OBLIGATIVE FUTURE as a Balkan phenomenon, to which the (Ukrainian) SYNTHETIC FUTURE is very likely to be related. It must be said that the DE-OBLIGATIVE form under study here is a better candidate to establish the link between the Balkans and Eastern Slavonic DE-OBLIGATIVES than the SYNTHETIC FUTURE, which is why I deal with this hypothesis in this section. In order to study the links between Eastern Slavonic and Balkan DE-OBLIGATIVES let us start from the only Slavonic languages in the Balkan group: Bulgarian and Macedonian.

7.4.2.3.1. Balkan-Slavonic links

According to Kuteva (1995) Bulgarian uses two main forms as a basis for the rest of the future forms: the form *šte* (a DE-VOLITIVE which I discuss in (§7.4.3.))²⁷⁹ for the positive forms; and for the negative forms the particles *da* (“an invariable conjunctive particle”) and *njama*, which actually is “the 3rd person singular present form of the verb *njamam* ‘have not’” (Kuteva 1995: 209). Even though the Bulgarian verbal system is more complex than that of West Polesian and BLM, one of the closest similarities is the possibility of expressing a counterfactual event grammatically.

“Bulgarian also has a more remote future perfect tense known as the “future in the past”, translated by ‘I was on the point of having done’” (Sussex & Cubberley 2006: 242).

So in this respect, the DE-OBLIGATIVE form in Bulgarian resembles the one in West Polesian (as well as the form that was used in Old Belarusian, *infra* (199)). In any case, besides the semantic basis and its possibility of expressing a counterfactual PAST event (for both IMPERFECTIVE and PERFECTIVE verbs (Kuteva 1995: 209-212)), the differences between the West Polesian and Bulgarian form are greater than the similarities. First, the polarity restrictions of Bulgarian do not apply to the West Polesian DE-OBLIGATIVE. Second, Bulgarian has gone further along the grammaticalisation path. As a result, Bulgarian *njama* is a semantically neutral form that can express both intentions and predictions; whereas West Polesian *majusj(a)* has a strong intentionality flavour. And third, Bulgarian *njama* does not currently inflect for PERSON and does not have a reflexive form, whereas

²⁷⁹ In Macedonian the cognate is *ke*, according to Dahl (2000b: 323).

the West Polesian DE-OBLIGATIVE is particularly salient (in the Slavonic family, at least) for having a reflexive form.

7.4.2.3.2. Other Balkan DE-OBLIGATIVES

We have seen in (§7.4.2.3.1.) that there are considerable differences between the Bulgarian and West Polesian DE-OBLIGATIVES. Nevertheless, we can still look for links with other non-Slavonic languages of the Balkans.

Dahl (2000b: 323) gives examples of DE-OBLIGATIVES in Gheg Albanian and Romanian. Danylenko's (2011: 164) comments suggest that the closest form (to Eastern Slavonic DE-OBLIGATIVES) in the Balkans is the Greek μέλλω which means 'to intend'. However, Danylenko (2011: 164-167, 177) also believes that the emergence of the Balkan de-obligative may have a Slavonic origin, although it would not be related to the forms that emerged in Middle Belarusian and Ukrainian:

“Historically, this instance of allegedly areal diffusion in the Balkans seems to be corroborated by the fact that, unlike all other Slavic languages, the P[eriphrastic] C[onstruction] with *iměti* was liberally used in Old Church Slavonic, presumably under the influence of Byzantine vernacular constructions with ‘to intend’ and ‘to want,’[...]

Identified sometimes as a Balkanism, a similar de-modal auxiliary was copiously attested in Old Church Slavonic, as well as later in the Russian recension of Church Slavonic due to the second South Slavic influence. In Ukrainian, however, the de-obligative paraphrase with the de-modal extension ‘to have to’ seems to be an independent development exhibiting no grammation of the auxiliary” (Danylenko 2011: 164, 177).

So far, I lack enough data and knowledge to decide whether the DE-OBLIGATIVE FUTURES in Eastern Slavonic (and particularly in West Polesian) are related or influenced by

analogous structures in the Balkans (as Dahl's (2000b) analysis suggests) or whether they are "indigenous developments" of the East (as Danylenko (2011) suggests). Nevertheless, it must be pointed out that none of the Balkan DE-OBLIGATIVES has a reflexive form (like WP *matysja*), which is a genuine innovation of Belarusian and West Polesian.

7.4.2.4. DE-OBLIGATIVES in a wider European context

As has been previously said, it is logical to assume that the Eastern Slavonic de-obligative forms (in particular, Belarusian and Rusyn) share the same etymology as the one in West Polesian. Yet the connection between the Balkan de-obligative and the one in Eastern Slavonic was not as obvious as it appeared at first glance. Thus, can the West Polesian (and Eastern Slavonic) DE-OBLIGATIVE be a loan or contact-induced by another European construction family? I address this question in this subsection.

So let us have a wider look into other European gram-families. Outside of the Balkans, the two best known de-obligative 'future construction families' are English *shall* (which earlier meant 'to owe'" (Bybee 1985: 194)) and the Romance SIMPLE FUTURE; e.g. [French] *aimer* + *ai* > *aimerai* 'I will love'. In comparison to West Polesian (and Eastern Slavonic), Romance DE-OBLIGATIVES are exclusively employed in the FUTURE TENSE. That is to say, they cannot be used as a FUTURE-IN-THE-PAST; because for those purposes, they tend to use the CONDITIONAL, which is also related to auxiliary *habere* (e.g. [French] *aimerais*) or most often an IMPERFECTIVE DE-ANDATIVE periphrasis. Besides, Bybee (1985: 194) also mentions Ukrainian (which I deduce, refers to the SYNTHETIC FUTURE form; *supra*) in the category of DE-OBLIGATIVE FUTURES.²⁸⁰

²⁸⁰ And outside of the European context, Bybee (1985: 194) also mentions Kru dialects.

In addition to these, within the European context, but totally unrelated to the Balkans or Slavonic, Basque has also a DE-OBLIGATIVE FUTURE TENSE, which is particularly used in familiar speech.²⁸¹ The future reading is deduced from the context. Compare (199) a. and (199) b.:

Basque

(199) a. [Context: a friend asking another friend before ordering.]

Kafe-a eska-tu **behar** (al) du-zu?
 coffee-DET.ABS.SG ask-INF need Q.PART have.3SG < 2SG
 ‘Are you going to order coffee? (lit. Do you have to order coffee?).’

b. [Context: Doctor talking to a patient.]

Ur gehiago edan **behar** du-zu.
 water.ABS more drink-INF need have.3SG < 2SG
 ‘You need to drink more water.’

Hence, there are DE-OBLIGATIVE FUTURES in different corners of Europe, which do not necessarily share much affiliation (if any). It has been briefly mentioned in (§7.3.3.), that according to Bybee’s (1985: 194) survey, modals denoting ‘obligation’ “tend to develop into *epistemic* markers” and “into *futures*, as in the case of English *shall*”. Ultan (1978: 114) had arrived at a similar conclusion a few years before Bybee’s (1985) survey:

“the semantic categories that tended most to give rise to future tenses were [...] quite a few modals, chiefly indicative of obligation, volition and uncertainty or unreality” (Ultan 1978: 114).

²⁸¹ Dahl (2000b: 325) also briefly mentions Basque as having two types of future constructions, although he hardly mentions it and does not include it with the other DE-OBLIGATIVE (mainly Balkan) FUTURE constructions.

Ultan (1978: 118) proposed that the link between modals and their common grammaticalisation into FUTURE TENSE markers cross-linguistically may be due “to the relative uncertainty inherent in both future events and most of the categories subsumed under the general heading of modals”.

That is to say, the Eastern Slavonic DE-OBLIGATIVE can be an independent development of the Balkan DE-OBLIGATIVE, given that modals frequently evolve into FUTURE TENSE constructions cross-linguistically.

7.4.2.5. Conclusion

In sum, besides the fact that all the mentioned languages have a similar or almost identical semantic base for their FUTURE TENSE constructions, there is no need to postulate a common origin or any genetic affiliation between them. As we have seen, DE-OBLIGATIVE FUTURES are fairly common cross-linguistically. Nonetheless, I also lack enough evidence to prove whether the grammaticalisation of Eastern Slavonic DE-OBLIGATIVES was motivated by a Balkan influence or not. Thus, my hypothesis so far is that the Eastern and Southern Slavonic DE-OBLIGATIVE FUTURE would have departed from a common material (the verb ‘to have’), that due to its semantics ‘naturally’ grammaticalised into a FUTURE TENSE gram. And, in this process, it is noteworthy that Belarusian and especially West Polesian have developed a DE-OBLIGATIVE FUTURE TENSE auxiliary based on a reflexive, which is nowhere else found in Europe, and that it expresses INTENTION rather than OBLIGATION.



Figure 12 Distribution of DE-OBLIGATIVE FUTURE in Europe. Sources (Bybee & Dahl 1989, Dahl 2000b, Danylenko 2011, Mazzitelli 2011, Whaley 2000).

7.4.3. The DE-VOLITIVE FUTURE

Closely related to the DE-OBLIGATIVE FUTURE (§7.4.2.), this is another construction in the spectrum of modality. There is evidence of the existence of DE-VOLITIVE and DE-OBLIGATIVE periphrases in older stages of Eastern and Common Slavonic (Danylenko 2011, Mackevič 1959, Whaley 2000). Thus, since this type of construction has a long tradition in Slavonic, it is not surprising to find it in other Slavonic varieties.

Starting from Eastern Slavonic, Mackevič (1959) vaguely mentions the existence of periphrastic constructions with the verb *xacec'* in Belarusian. However, he does not provide any examples, nor does he say in which dialects this form is used (certainly, it is not part of the current standard BLM, so he may be referring to West Polesian or South Western Belarusian, in general). According to Danylenko (2011: 154), Transcarpathian (i.e. Rusyn) and Eastern Ukrainian also have DE-VOLITIVE periphrastic constructions based on the verb *xotity*, although they have not fully grammaticalised the verb into a FUTURE TENSE auxiliary yet.²⁸² In this respect, West Polesian is probably not the only member of the Eastern Slavonic to have preserved a DE-VOLITIVE form. Nonetheless, it has gone further in the path of grammaticalisation than any other member and it has made the biggest switch in the semantic content of the base: from volition to obligation.²⁸³

Moving further south, Bosnian-Croatian-Montenegrin-Serbian (hereafter BCMS) forms its FUTURE TENSE using the inflected form of the verb *hteti* / *htjeti* 'want' as an auxiliary, which is a cognate of West Polesian *xotiti*. When it is in a positive sentence, the auxiliary can appear either as the whole form (*hóču/hóčeš/etc.* + INFINITIVE) or as a shortened form (*ču/češ/etc.* + INFINITIVE). However, when the verb is negated, only the NEG *ne* + the short form is possible (*nè ču* or *nèču /ne češ* or *nečeš* /etc. + INFINITIVE; but * *ne hóču/etc.* + INFINITIVE) (Alexander 2006: 143-144).

²⁸² Unfortunately, Danylenko (2011) does not provide more details or examples; nor does Pugh's (2009) Rusyn grammar make any allusion to it.

²⁸³ HSBM (vol.37) does not record any deontic meaning of the verb *xotiti*. Lena V. Levancèvič (p.c.) shared with me that other West Polesian sub-varieties (as in Malaryta and Southern Byarozhy) also use *xotiti* with a deontic meaning. She considers it to be a specificity of West Polesian/Zaharoddzian.

Gvozdanović (1995: 186, 189, 190) says that BCMS FUTURE TENSE (regardless of ASPECT) is used:

“(a) for a future event, as in (200) a. ;

(b) for a necessary or desired event, as in (200) b.”

Gvozdanović (1995: 186) shows the following examples of BCMS FUTURES:

BCMS (Gvozdanović 1995: 186)

(200) a. Sutra ć-u ići u grad.
tomorrow will.PRS-1SG go.INF in/to city.ACC.SG

‘Tomorrow I shall go to the city.’

b. Sada ć-eš ići po to.
now will.PRS-2SG go.INF for this.ACC.SG

‘Now you will go to fetch it.’

As we can appreciate from Gvozdanović's (1995: 186) comments (particularly, see translation for (200) b., this DE-VOLITIVE construction shares with its West Polesian cognate part of the NECESSITY meaning (which is a further innovation, given the etymology of the gram). Moreover, both constructions combine with PERFECTIVE as well as IMPERFECTIVE infinitives, which is an uncommon feature with regards to other constructions in WP (and Eastern Slavonic).²⁸⁴ However, the usage of this form is far more limited in West Polesian than in BCMS. BCMS allows the use of this form as a semantically neutral FUTURE form (i.e. for both intentions and predictions); whereas West Polesian DE-VOLITIVE implies the existence of a need to be addressed. In exchange,

²⁸⁴ According to Gvozdanović (1995: 186-187, 192) the BCMS and Slovene ‘future perfects’ can be used with either PERFECTIVE or IMPERFECTIVE verbs (though they are based on the copular form).

West Polesian, having a morphologically richer repertoire of constructions, most frequently uses the COPULAR form in such a context (i.e. a semantically neutral FUTURE or a prediction).

Bulgarian/Macedonian also has a future tense auxiliary which was etymologically a volitional verb, *šta* ‘will/want’, (cognate and related of BCMS *hteti/htjeti*).²⁸⁵ Nonetheless, the auxiliary is almost exclusively used in positive sentences (with the vast majority of negatives it uses the DE-OBLIGATIVE). The FUTURE (SIMPLE) is formed by the auxiliary *šte/ke*, etymologically the 3SG form of the verb *šta* (although the auxiliary itself does not inflect for PERSON), and the PRESENT TENSE form of the main verb (Kuteva 1995: 209).

Bulgarian (Kuteva 1995: 209)

- (201) *Šte* *dadeš*
 will/want(INVAR) give.PFV.2SG.PRS
 ‘You will give’.

The Bulgarian DE-VOLITIVE FUTURE form is versatile with regard to TENSE and is the base of the FUTURE SIMPLE; FUTURE PERFECT; FUTURE-IN-THE-PAST and FUTURE PERFECT-IN-THE-PAST forms (Kuteva 1995: 209-212).²⁸⁶ Even the FUTURE SIMPLE form (as in (201)), can express “an activity which will take place [...] or will be taking place [...] after the moment of speech” or “a repeated or habitual activity either in the present or in the past” (Kuteva

²⁸⁵ According to Andrew Spencer (p.c.) in contemporary BCMS that verb has been replaced by *iskam*.

²⁸⁶ Note that in the FUTURE-IN-THE-PAST and the FUTURE-PERFECT-IN-THE-PAST the auxiliaries inflect for PERSON as well as TENSE.

1995: 209).²⁸⁷ By contrast, the West Polesian DE-VOLITIVE is very restricted in TENSE (in contrast with other West Polesian FUTURE-TENSE constructions also discussed here): it can only be used in the FUTURE TENSE (by inflecting the auxiliary in the PRESENT TENSE), as in any other form it functions as a plain volitive verb (see (§7.2.3.2.)).

Elsewhere in the Balkans, Dahl (2000b: 323) mentions further languages with DE-VOLITIVES. On the one hand Modern Greek and Gheg Albanian which in the same vein as Bulgarian/Macedonian, have an uninflectable form of the verb ‘want’ as the semantic base of their FUTURE TENSE auxiliary (Duchet 1995: 257, Hedin 1995: 233).²⁸⁸ And on the other hand Romanian, like BCMS, has an inflected auxiliary meaning ‘want’. Having said this, we should not be surprised to find more DE-VOLITIVE FUTURE constructions in the world’s languages. Since volition and intentionality are semantically very closely related, volitive verbs often develop into FUTURE TENSE constructions (Bybee 1985, Bybee & Dahl 1989).

Outside of the Balkan context, Dahl (2000b: 322) finds a second DE-VOLITIVE FUTURE construction family within the Germanic languages of the North Sea: Danish, Faroese and Norwegian Bokmål (*vil*); Frisian (*wal*); and perhaps also Yiddish (*vel*). The English FUTURE TENSE construction *will* is also originally a volitive verb.²⁸⁹ According to Dahl

²⁸⁷ Andrew Spencer (p.c.) pointed out to me that English *will* futures can also express generic habitual events or past events; e.g. “*By 1811 Napoleon will become master of half of Europe; by 1815 he will be defeated and in exile*”. He also remarked that in Bulgarian RENARRATED or EVIDENTIAL MOOD you also encounter FUTURE TENSES.

²⁸⁸ In Modern Greek (*tha*) and Bulgarian/Macedonian (*šte/ke*) the verb of the auxiliary has not survived as an independent volitional verb to our days.

²⁸⁹ According to OED (2018) in Contemporary English the verb *will* can mean “[t]o wish, desire; sometimes with implication of intention” although it classifies it as obscure, rare and archaic.

(2000b: 322), it is the most grammaticalised FUTURE TENSE construction of all the languages in this construction family.

Other than the semantics of the auxiliary chosen as a base for the FUTURE TENSE, it is unlikely that there was an areal influence of the North Sea/Germanic *will* (and its cognates) in the development of West Polesian *xotfu*. First, historically it is hard to prove (especially given that the closest Germanic language that could have been affected, High German, did not develop it). And second, there is no phonological correspondence between West Polesian *xotfu* and Germanic *will* (and cognates); whilst *xotfu* can be clearly related to the Southern Slavonic forms like *(hó)ču*.

Thus, where does the West Polesian DE-VOLITIVE FUTURE stand with regards to the Balkan (Southern Slavonic) DE-VOLITIVE? On the one hand, we can hypothesise that the West Polesian DE-VOLITIVE is a loan or calque of Southern Slavonic. There are a few other features in West Polesian which look like Southern Slavonic (mostly Serbian; e.g. VOCATIVE MASCULINES in *-u*; or the DE-OBLIGATIVE FUTURES).²⁹⁰ But on the other hand, it is more plausible to believe that the DE-VOLITIVE FUTURE in West Polesian can be an independent development departing from a common base that has only grammaticalised in a few modern Slavonic varieties. According to Danylenko (2011: 162), in Common Slavonic “the verbs *iměti* ‘to have’ and *xotěti* / *xotěti* ‘to want’ were used in such P[eriphrastic] C[onstruction]s as de-modal extensions rather than auxiliaries.” In any

²⁹⁰ And in fact some speakers have told me (although I lack data to prove this) that there was a wave of migration from the south (mostly around contemporary Serbia) to the lands of West Polesie around the XV century (which would be reflected in the common surnames). But again, I am unaware of any academic sources on this topic.

case, the final output of the contemporary semantics of *xotfu* diverges from the rest of the construction families (and from typological expectations), which is a remarkable phenomenon that deserves further study in its own right.

Against the general typological expectation (in Figure 13 represented by English), the West Polesian DE-VOLITIVE and DE-OBLIGATIVE constructions have both retained some MODAL value, but they have inverted their original semantic functions. Southern Slavonic (particularly Bulgarian and Macedonian) retains both (cognate) constructions, but it has generalised their use; that is to say, it has gone further in grammaticalisation (especially if it is understood as ‘semantic bleaching’). By contrast, synchronically the West Polesian DE-VOLITIVE form has ended up expressing necessity, whereas the de-obligative is used to indicate intention and desire.

	English	Southern Slavonic	West Polesian
Intention, desire	<i>will</i>	(BCMS) (<i>hó</i>)ču ‘want’	<i>majusj(a)</i> ‘have’
Prediction	<i>will</i>	(BG/MKD) + <i>šte/ke</i> ‘want’	(COPULAR, SYNTHETIC, perfective or DE-VENTIVE futures)
Obligation	<i>shall</i>	– [<i>nja</i>] <i>ma</i> ‘have [not]’	<i>xotfu</i> ‘want’

Figure 13 Examples of uses of future tense constructions

How did West Polesian end up in this situation? That is to say, how did both constructions end up inverting their original meanings? I lack sufficient data to investigate this, but it is likely that this overlapping of meanings is merely accidental; i.e. both forms were in the process of desemanticisation/generalisation, and they both ended up finding ‘an empty nest’ at different times in history. From a more simplistic approach, this semantic

‘dissonance’ can be also the result of a saturated inventory of FUTURE TENSE constructions in West Polesian, which may have ‘forced’ certain members to specialise over time. Unfortunately, I am not in a position to provide any meaningful answers, so I hope further research on this topic will shed more light on this question.²⁹¹

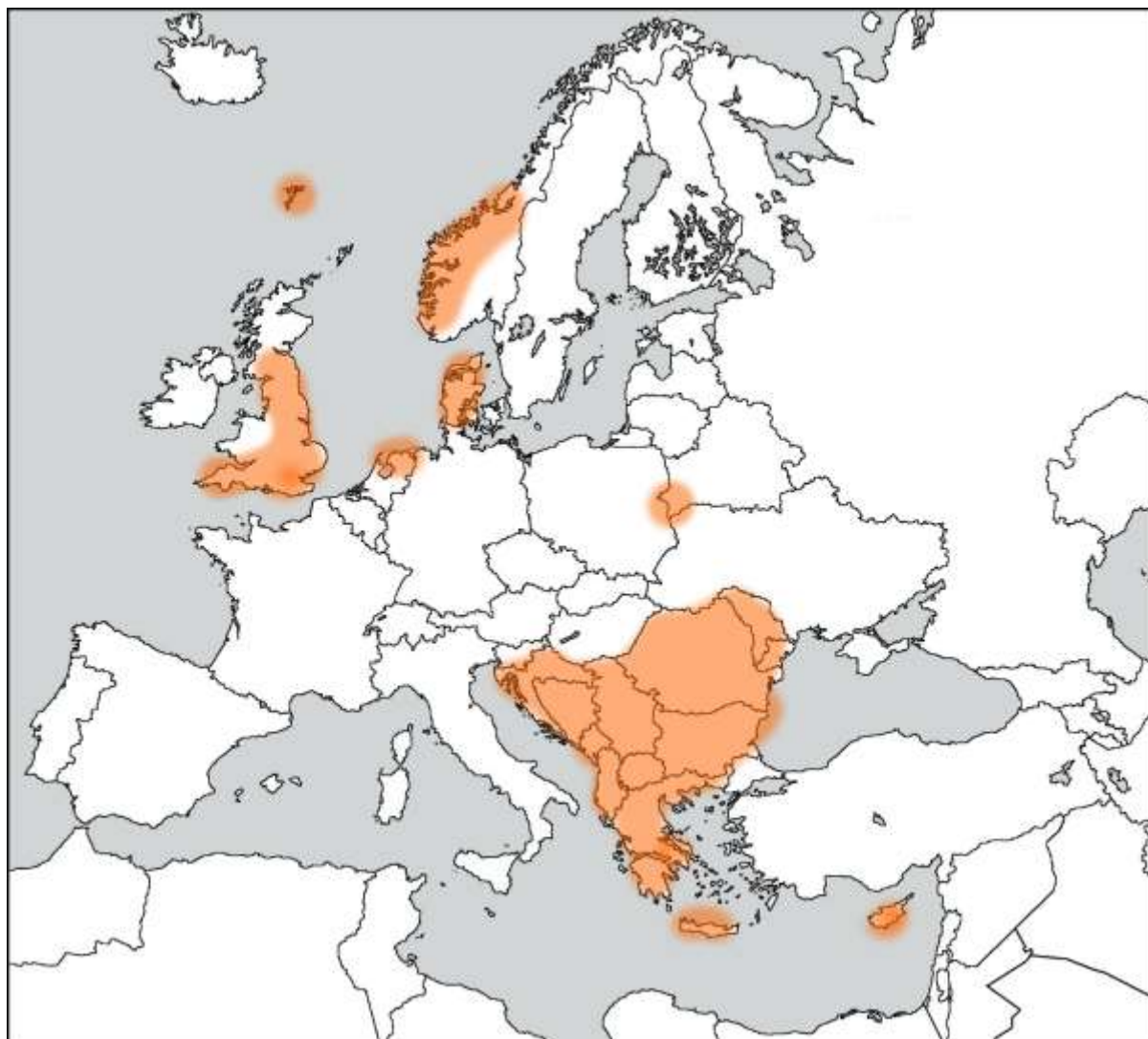


Figure 14 Distribution of DE-VOLITIVE FUTURE in Europe.

Sources (Dahl 2000b, Danylenko 2011, Duchet 1995, Gvozdanović 1995, Kuteva 1995, Mackevič 1959).

²⁹¹ According to Andrew Spencer (p.c.), English also had a similar type of ‘inversion’. The de-volitive **will** used to mean plain volition (as in *as you will*), which then became a FUTURE AUXILIARY. Conversely, **want** used to mean necessity or need (as in *for they shall want for nothing*), but then filled the slot vacated by **will**.

7.4.4. The DE-VENITIVE future tense

In (§7.2.4.) we have seen that the auxiliary verb *statı* is more often the head of inchoative periphrastic constructions (particularly in the PAST TENSE), although it can sometimes act as a ‘neutral’ FUTURE TENSE construction (i.e. lacking additional semantic nuances) in West Polesian. In this sub-section I explore other DE-VENITIVE future constructions in Europe and their relation to the West Polesian DE-VENITIVE.

7.4.4.1. Slavonic links

According to Whaley (2000: 142), the DE-VENITIVE type of FUTURE TENSE construction exists in all modern Eastern Slavonic languages. Nonetheless, I struggled to find it in descriptions of Contemporary Standard Belarusian (BLM). For example, Mackevič (1959: 199-200) briefly describes the existence of such periphrastic forms in Standard Belarusian (BLM), but none of the examples that he provides have proper FUTURE TENSE forms.²⁹²

Notwithstanding the lack of attention to (or even mention of) it in most contemporary grammars, Contemporary Standard Russian (CSR) also has a DE-VENITIVE future. See example (202):

CSR (Naši deti: Podrostki (2004) In Russian National Corpus (2017))

- (202) Bo-ju-s', čto kolledž ego rasslab-it i
fear-1SG-REFL COMP college-NOM.SG 3SG.ACC.M relax.PRF-3SG and
on ne stan-et dal'she uči-t'-sja.
3SG.NOM.M NEG become.PRF-3SG further study-INF-REFL
'I'm afraid college will make him too relaxed and that he will not study
any further.'

²⁹² According to Lena V. Levancëvič (p.c.), DE-VENITIVE forms (i.e. *stanu* + INF) are dispreferred in BLM.

Whaley (2000: 61-62) says that it is possible to find the DE-VENITIVE form in CSR literary texts, although it is often “an imitation of folk language”. In fact, according to her, the DE-VENITIVE construction is a lot more robust, to the point of even overtaking the COPULAR form, in certain dialects of Russian, particularly in the North (Vologda) (*ibid*).

7.4.4.2. Properties

Whaley (2000: 153) says that the verb *stat’/staty/stac’* ‘to become; to begin’ has multiple meanings in Eastern Slavonic. The fact that the *Historical Dictionary of Belarusian* (HSBM) has eighteen definitions for it gives good evidence of this (see more in (§7.4.4.3.)). Nevertheless, in Eastern Slavonic this verb can also be used as a FUTURE TENSE auxiliary (with greater frequency and sociolinguistic acceptance in some varieties than in others) “subject to the same colligability constraints that one typically finds for the be-future [the COPULAR future]” (Whaley 2000: 153).

Eastern Slavonic DE-VENITIVES are characterised by having an auxiliary that has a DEFECTIVE ASPECT (Whaley 2000: 141). And, as we have seen in (§7.2.4.) this also applies to West Polesian DE-VENITIVES. In all the Eastern Slavonic sub-family the DE-VENITIVE FUTURE can only be used with IMPERFECTIVE infinitives, even though the auxiliary verb itself is in the PERFECTIVE FUTURE and lacks an IMPERFECTIVE counterpart. In some contexts the inchoative meaning is stronger than the ‘pure’ temporal (i.e. FUTURE) value, denoting “transition to the realisation of the action (*stanu čytač*) [‘I will start to read’]” (Mackevič 1959: 199-200) [My translation]. The auxiliary is more commonly used in the PAST TENSE, with a stronger inchoative meaning (of the type of

future-in-the-past). Yet, the form we are most interested in is when the auxiliary appears in the (PERFECTIVE) FUTURE.²⁹³

7.4.4.3. Chronology

Whaley (2000: 142) summarises the causes that led the verb *stat'* (with all its variations) to become a FUTURE TENSE auxiliary:

“First, there is the evidence that the verb *stati* has become polysemous over time, with *stati*¹ retaining change-of-state meaning and *stati*² developing into an inceptive phase verb. Proposing such a meaning-shift for **bqdaq* allows one to explain the colligability of the modern auxiliary. *Stati*² became a verb which takes infinitive complements, as has **bqdaq*. [...] Second, *stati*² also (or perhaps only subsequently) underwent at least partial grammaticalization into a future auxiliary, with inceptive meaning giving way to future meaning” (Whaley 2000: 142).

The Historical Dictionary of Belarusian (HSBM vol. 32) records up to eighteen different meanings for the verb *stati* (or its variations *statisja*, *sstatise*, *statise*, *stat'sje*, *stat'sja*). Only the tenth definition briefly documents the use of *stati* denoting ‘to begin’ in older stages of Belarusian, as in (203) a. and b.; and the thirteenth definition documents *stati* meaning ‘to happen’ (which are closer to its contemporary meanings in West Polesian), as in (203) c.

²⁹³ This has some presence in different Russian varieties, but it is certainly marginal in BLM (if it is present at all).

Old Belarusian

- (203) a. [Tr. Hist., 71b] see further references in (HSBM vol. 32)

коли	впрывадили	его	до	мѣста,	стали
koli	uprovadi-l-i	eho	do	měst-a,	sta-l-i
when	lead.PRF-PST-PL	3SG.ACC.M	to	city-GEN.SG	become-PST-PL

всѣ	ѣсти	и	питы	и	играти
vs-i	es-ti	i	pi-ti	i	ihra-ti
all-NOM.PL	eat-INF	and	drink-INF	and	play-INF

‘When they took him to the city, they all started to eat, drink and play.’

- b. [Byx., 524] see further references in (HSBM vol. 32)

Nemc-y	Prusow-e	y	Liflant-y	sta-l-i
German-NOM.PL	Prussian-NOM.PL	and	Liflant-NOM.PL	become-PST-PL

z	woysk-i	na	nich	zbira-ty
from	army-GEN.SG	to	3PL.ACC	choose.PRF-INF

‘(Germanic) Prussians and Liflants started to select [people] from the army.’ [Uncertain meaning]

- c. [Čèccja, 263] see further references in (HSBM vol. 32)

стал	страхъ	великъ	во	всѣмъ	городѣ
sta-l	straxъ	velik-ъ	vo	vs-ěmъ	horod-ě
begin-PST.M.SG	fear(M).NOM.SG	big-NOM.SG	in	all-LOC.SG	city-LOC.SG

‘A great fear invaded the whole city (lit. emerged in the whole city)’.

As we can guess from the examples provided for the uses of the verb *stati* the INCHOATIVE/DE-VENITIVE construction was used primarily in the PAST TENSE in the earliest stages of Belarusian/Eastern Slavonic.²⁹⁴

²⁹⁴ It must be said that none of the resources consulted provide much evidence of the use of the verb *stac* ‘as an auxiliary (even for inchoatives) either in Contemporary or in older stages of Belarusian.

We know that the form *stanu* was in use by the sixteenth century in Russian, and that during the seventeenth century both the COPULAR and the DE-VENITIVE constructions were frequently used (Whaley 2000: 69). Already by the end of the seventeenth century, both constructions started to be described in grammars, although no meaningful explanations of any differences between them were given for a century (Whaley 2000: 70-71). That is to say, both constructions were used on a regular basis, more or less as synonyms.

“Whereas Lomonosov only hinted at the idea that *stanu* and *budu* were not employed as functional equals, now one finds the perspective that *budu* is the only “pure” future auxiliary. Such a position is found in Buslaev’s (1959) grammar, originally published in 1881” (Whaley 2000: 71).

In fact, according to Whaley (2000), there is evidence of a prolonged co-existence of multiple FUTURE TENSE constructions in Russian which lasted for centuries (and which is partly attested in Buslaev’s (1959) grammar).

7.4.4.4. Links to other European forms

According to (Dahl 2000d: 357) the become type of Future Time Reference expressions (or “extended uses of ‘become’ verbs) exist in every Germanic language (English being the exception). In fact, DE-VENITIVES have been attested in Gothic: “[i]n grammars of Gothic, the use of the verb *wairþan* ‘to become’ for ‘will be’ is regularly mentioned [...]” (Dahl 2000d: 357). It is also interesting to note that according to Dahl (2000b: 322-323) in Old High German ‘become’ had two values: inchoative and modal, which match with what we have seen for West Polesian (and Eastern Slavonic).

In addition, Dahl (2000d) provides examples of Hungarian and Finnish where the verb ‘become’ is also used with some prospective/temporal meaning. And he points out that the Estonian construction is “regarded as a calque of the German *werden* future” (2000b: 323-324). However, regarding all these Germanic and Finno-Ugric DE-VENTIVES (or “‘become’-type constructions”) he says:

“[...] what we are here dealing with is, in principle, a special use of verbs of becoming rather than a future tense marker. But there is still a clear link to future time reference. [...] It is thus possible that even if the Germanic and Baltic-Finnic verbs of becoming cannot be regarded as future copulas, the extensions of their use that we can observe represent the first step in such a grammaticalization path.”(Dahl 2000d: 353, 359-360)

That is to say, the become-type (or DE-VENTIVE) constructions are not fully grammaticalised FUTURE TENSE forms in Contemporary Germanic and Baltic-Finnic (i.e. Northern European) languages (except for Yiddish, where it has become fully SYNTHETIC (Dahl 2000d)). In fact, (Dahl 2000d: 359-360) sees a correlation between the frequency of ‘become’ and the “futureless areas” (Dahl 2000b) in Northern Europe, Slavonic being the exception. Nonetheless, the DE-VENTIVE future is also in the process of decay in Eastern Slavonic in comparison with older stages where it was in a much stronger competition with the COPULAR form. And this decay is noticeable from Standard Contemporary Russian (the best documented) to West Polesian.

On the one hand, according to Bybee (1985), Whaley (2000) and Ultan (1978), the semantics of change-of-state verbs like ‘become’ make it a very good candidate to

grammaticalise as a FUTURE TENSE marker (cross-linguistically).²⁹⁵ So while the similarities between the different FUTURE TENSE constructions may be striking, we should not be too quick to jump to conclusions about a common origin.²⁹⁶

On the other hand, it is a matter of debate whether the Slavonic IMPERFECTIVE FUTURE was a borrowing from German *werden* ‘become’, or whether German borrowed from it Czech.²⁹⁷ According to Whaley (2000: 98), there is evidence of Czech-German linguistic contact, which may point to an areal development. Nonetheless, she believes that there is a greater chance that German borrowed it from Slavonic, given that the latter is more complex in having all the aspect restrictions (see more in (§7.4.6.)).

As far as I have been able to conclude, the development of Future Time Reference ‘become’-type time expressions is an areal development. Northern European languages have taken advantage of a type of verb (change-of-state) which is typologically prone to turn into a FUTURE TENSE auxiliary. Eastern Slavonic and particularly Yiddish and High German have gone further in grammaticalisation of the DE-VENITIVE verb as a ‘neutral’ FUTURE TENSE construction; whilst other Circum-Baltic languages, particularly Finnish and Estonian, have not fully grammaticalised the DE-VENITIVE form. This reluctance in the development of a FUTURE TENSE construction can be explained by the fact that Estonian and Finnish lack a systematic (overt) grammatical differentiation between PRESENT and FUTURE TENSE. By

²⁹⁵ “In general, the semantic shifts and grammaticalization of the verb *stati*² [e.g. in its de-venitive meaning] in East Slavic lend compelling support to the hypothesis that change-of-state verbs can develop into an imperfective future auxiliary” (Whaley 2000: 143).

²⁹⁶ The phonological correspondence between the Germanic *werden* (and derived) and the Slavonic *stanu* are obviously null. If there was any kind of relation between them, it could only be calque.

²⁹⁷ Most of the debate is centred on the copular form, but with more reason the hypotheses and conclusions can be applied to the DE-VENITIVE form.

contrast, we have seen that in Eastern Slavonic multiple future constructions have existed through history, so it may have been easier to integrate a new member.

In any case, it may be that the DE-VENTIVE future form presents certain semantic or structural problems that hinder the process of fully grammaticalising as a FUTURE TENSE marker for most members of the ‘construction family (or families)’. According to Dahl (2000d: 354) “[s]entences expressing location present special problems” in German and Swedish; and this is also the case for West Polesian and Eastern Slavonic.²⁹⁸ Dahl (2000d: 354) says that “[t]his is fairly natural, since Swedish *bli* and German *werden* do not in general function as inchoatives in locative constructions”.

7.4.4.5. Summary

Summing up, the DE-VENTIVE future construction can be found throughout Eastern Slavonic (although very marginally in Belarusian). For the vast majority of Eastern Slavonic varieties, the COPULAR form has nowadays overtaken this form, which is now considered “dialectal”. Nevertheless, there is historical evidence of it being far more robust and in much closer competition with the COPULAR form in the PAST. This construction is also present in Germanic and Finno-Ugric (European) families, although with a few exceptions (Yiddish and High German) it is more advanced in its grammaticalisation (as a future marker) in Eastern Slavonic. The exact origin of the form is unclear; hence, it is reasonable to postulate that it could be an areal development,

²⁹⁸ Unless the verb *stat’* (and derived) is being used as ‘to stay’ (e.g. at home), and curiously enough, Swedish and German can also use their ‘become’ verbs in that sense (Dahl 2000d). Note also that according to Dahl (2000d) Hungarian shows a lot more flexibility with this type of constructions, although he also admits that “even in Hungarian, the treatment of ‘become’ as a future copula is somewhat problematic” (2000d: 357).

admitting that the semantics of the base is a common source of FUTURE TENSES cross-linguistically. DE-VENTIVE constructions, however, present several semantic restrictions in most European languages (e.g. location), which may have an effect in grammaticalisation; i.e. it would put this construction in a less favourable position (with respect to other FUTURE TENSE constructions). Thus, this may have contributed to the victory of the COPULAR future over the DE-VENTIVE in Eastern Slavonic.



Figure 15 Distribution of DE-VENTIVE FUTURE in Europe.

Sources (Dahl 2000b, 2000d, Whaley 2000)

7.4.5. PERFECTIVE FUTURE (or PRESENT)

7.4.5.1. Cross-Slavonic overview

The Slavonic PERFECTIVE FUTURE (or PRESENT) exists in every Slavonic language, although in Bulgarian and Macedonian, which have a more elaborated verbal system, there are several forms (Kuteva 1995, Sussex & Cubberley 2006). Thus, many linguists (like Dahl (2000b: 323, 326)) consider this to be a purely Slavonic development (or gram), which also covers West Polesian.

The functions and uses of the PERFECTIVE FUTURE are very homogeneous across the Slavonic family. For example, what Schuster-Šewc (1999: 169) says about the Upper Sorbian PERFECTIVE FUTURE, as opposed to “the periphrastic future” (i.e. the COPULAR FUTURE) can easily be applied to any Eastern Slavonic variety (including West Polesian):

“The future tense expressed by the non-past form of a perfective verb differs from the periphrastic future in that the former denotes a verbal action whose *completion* is envisioned by the speaker as being in the future; the periphrastic future, in contrast, does not specify such completion [...]. The non-past form of a perfective verb is ambiguous. It can express a non-continuous (non-eventive) present as well as completed future action” (Schuster-Šewc 1999).

Thus, in comparison with the rest of the Slavonic family, West Polesian does not stand out by virtue of either the function or the morphology of this construction. Yet, since West Polesian has more constructions that combine with PERFECTIVE verbs (the DE-OBLIGATIVE and the DE-VOLITIVE forms), it differs from the majority in that the use of the PERFECTIVE FUTURE is not always obligatory (admitting that the alternatives bring an additional semantic load).

7.4.5.2. Controversies regarding the PERFECTIVE FUTURE

I have already introduced the debate in (§7.1.2.), which becomes particularly relevant at this point. Some linguists deny the existence of the FUTURE as a value of TENSE (at least in Eastern and Western Slavonic). For example, Vater (1995: 153) says that Polish only has two tenses “present and past, both occurring with the imperfective and perfective aspect respectively.” However, the PERFECTIVE FUTURE or PRESENT can be rarely employed to make reference to an event happening in the moment of speech in Slavonic languages (i.e. it cannot be used as a ‘straightforward’ PRESENT) (Sussex & Cubberley 2006). So, how should we analyse the PERFECTIVE FUTURE? According to Vater (1995: 156), the key is in the semantic restrictions of PERFECTIVE ASPECT:

“The perfective present tense cannot express presentness any more, since the description of an event going on during the time of speech is incompatible with the meaning of the perfective aspect (in all its interpretations [...]). This is the consequence of the semantics of the perfective aspect rather than of the semantics of the present tense” (Vater 1995: 156).

I lack enough data and knowledge in formal semantics to argue against such claims. Nevertheless, I will follow the mainstream opinion of Slavists and still decide to analyse it as an expression of the FUTURE TENSE value which combines with PERFECTIVE aspect. Now, since the PERFECTIVE FUTURE is a purely SYNTHETIC form some Slavists consider it the only proper FUTURE TENSE form (e.g. Gvozdanović 1994).²⁹⁹ However,

²⁹⁹ “Concerning the Russian tense, we have seen above that the system is based on two simple tenses: the past and the present tense. The fact that a tense may be used with reference to a temporal level different from its basic or prototypical one is referred to in the literature by the term ‘transposition’ [...]” (Gvozdanović 1994: 193).

such a claim is not unproblematic. First, Bybee (1985: 101, 146) stresses that the value of PERFECTIVE in BCMS (and this can equally be applied to Eastern and Western Slavonic) is realised more derivationally than inflectionally (and thus, it is less grammaticalised). The ways of deriving the PERFECTIVE stems (from the IMPERFECTIVES) show certain regularities or tendencies, but they are not fully predictable (and consequently, lexically specified).³⁰⁰ Second, in Slavonic languages often the prefixes added to derive the perfective form bring additional semantic values to the verbal form (other than just purely aspectual/temporal). And in some cases, there are multiple possibilities (suffixes) for deriving the perfective stem (each with an additional nuance), so the correspondence between the IMPERFECTIVE and the PERFECTIVE is not straightforward. By contrast, the COPULAR FUTURE TENSE is realised regularly in every Slavonic variety.³⁰¹ And third, as I have already pointed out in (§7.1.2.), we find that the FUTURE TENSE is very frequently realised periphrastically in the world's languages (Bybee 1985).

7.4.5.3. Summary

In sum, West Polesian PERFECTIVE FUTURE (or present) can be mapped within the Slavonic PERFECTIVE FUTURE construction family. As with the rest of the construction family, its realisation relies heavily on the formation of the perfective stem, which involves a high element of derivation and thus less predictability. The only

³⁰⁰ The most common strategy is prefixation, but stress shifts or vowel alternations of the stem (umlauts and ablauts), or even a combination of several strategies are also used in West Polesian in order to derive a perfective from an IMPERFECTIVE.

³⁰¹ Some verbs, particularly verbs of motion, may have restrictions, but these are probably due to their semantics rather than purely lexical reasons. Moreover, they only comprise a small group.

particularity that makes West Polesian stand out from the construction family is the PERFECTIVE FUTURE is not always obligatory with perfective verbal bases, whilst for most Slavonic languages it is.



Figure 16 Distribution of PERFECTIVE FUTURE (Dahl 2000b, Sussex & Cubberley 2006).

7.4.6. Slavonic COPULAR constructions

At first glance, there is a COPULAR FUTURE TENSE in every contemporary Slavonic language except for Bulgarian and Macedonian. For this reason, Dahl (2000b: 324) includes all of

them under the “family” of “Slav[on]ic copular constructions”. The West Polesian COPULAR FUTURE TENSE displays similarities with other COPULAR constructions across Slavonic varieties. However, the relation between the different COPULAR constructions is not uncomplicated. They have a heterogeneous realisation, in terms of the verbal ASPECT they can combine with (i.e. only IMPERFECTIVES, or BI-ASPECTUAL, like in West Southern-Slavonic and informal Upper Sorbian (Schuster-Šewc 1999)) or the complement/non-finite verb (infinitives and/or participles).³⁰² Moreover, some are attested earlier than others in the literature. It is important to remember that the COPULAR FUTURE TENSE is not attested in Old Church Slavonic (OCS). Given that OCS is the earliest written record we have of a Slavonic language, to Whaley (2000: 25) that lack of records demonstrates that the development of the COPULAR FUTURE happened after the split into individual Slavonic languages. In fact, Whaley (2000: 152) warns against historico-typological approaches that treat all the different COPULAR FUTURE constructions (‘be-futures’ in her terminology) in Slavonic as being the same (and having a common origin):

“[...] is in fact a set of constructions with diverse meanings, uses, and paths of development in the individual Slavic languages. The major flaw of previous theories is their failure to account for this diversity among be-future constructions; they tend to view the *be*-future as a single, unified type of construction for which a single point of origin is assumed” (Whaley 2000: 152).

Whaley’s (2000: 152-153) position on the diversity of COPULAR constructions is that they developed independently. On the one hand, they took material from the

³⁰² “Unlike most other Slavic languages, the Upper Sorbian popular language forms a periphrastic future with perfective verbs: *budu wohladać* ‘I shall catch sight of’, *budu namakać* ‘I shall find’, [...]. Such future-tense formation runs counter to the norms of the Upper Sorbian literary language” (Schuster-Šewc 1999: 169).

PERFECTIVE constructions. On the other hand, this was motivated by the semantics of the form **bqdaq* ‘to be’ (inherited by all Slavonic languages from Common Slavonic), which can also be interpreted as a DE-VENITIVE type of verb (i.e. a change of state verb). And, as we have already seen in (§7.4.4.), DE-VENITIVE verbs frequently evolve into FUTURE TENSE REFERENCE markers.

Having said this, the West Polesian COPULAR FUTURE TENSE has the same morphological structure and semantics as the rest of the Eastern Slavonic family (except for a few Belarusian and Ukrainian dialects on the Polish border, where the participle is also allowed) (Lomtev 1956, Pugh 2009, Whaley 2000). Thus, it is reasonable to claim a connection between all the varieties of the Eastern Slavonic group (including Rusyn).

According to Whaley (2000: 64), the earliest attested instance of a COPULAR FUTURE in Eastern Slavonic is from Article 99 of the “Russkaja Pravda” law code, whose earliest edition is dated circa 1280. However, the validity of that token as a genuine future form is highly disputed. The next instances of a COPULAR FUTURE in Eastern Slavonic are from the Great Duchy of Lithuania (from primarily the area of contemporary Belarus) dating to 1375 and 1388; but it will not be until the period of Kazimierz Jagiellończyk (second half of the fifteenth century) that its use becomes more frequent in chancery texts (Jankoŭski 1989: 230, Whaley 2000: 54-56). As already mentioned, in older stages of Eastern Slavonic different periphrastic constructions existed, so the COPULAR form was only one of them, very marginal at its start. However, in CSR and BLM this form has overtaken all the IMPERFECTIVE periphrastic forms and this is also happening in certain varieties of West Polesian. Only Ukrainian, some South-western Belarusian and some West Polesian

varieties show resilience to the assimilation. In any case, the functions of this form are virtually identical across all Eastern Slavonic varieties.

In (§7.4.4.) we have seen that many Germanic languages had a DE-VENTIVE FUTURE. Because of the similarities between the High German *werden* ‘become’ and the Slavonic COPULAR constructions some authors have argued that Slavonic (more specifically, Czech) borrowed its COPULAR construction from Germanic and vice versa (read more on this debate in (Dahl 2000b, Leiss 1985, Whaley 2000)). Whaley (2000: 153) favours Leiss’s (1985) hypothesis: “that German borrowed its change-of-state future from Slavic since the construction is aspectually motivated in Slavic but not in German”.³⁰³ Although she also recognises that DE-VENTIVE verbs are a fairly common source of FUTURE TENSE constructions cross-linguistically (as already pointed out in (§7.4.4.)), without the need for borrowing or calquing from other languages. So admitting her limitations and the lack of more data, she eventually proposes “that the North Slavic participial be-future is, in fact, an areal development” (Whaley 2000: 145-146).

In sum, there are different COPULAR FUTURE constructions in Contemporary Slavonic languages. Because of their morphological heterogeneity and differences in chronology, it does not seem appropriate to classify all of them as a single group, although Eastern Slavonic (including West Polesian) is quite homogeneous in this respect. The best explanation so far for the relation between Eastern Slavonic COPULAR FUTURES and the rest of the Slavic family is that they all departed from a common

³⁰³ Whilst she also admits that the fact that West Southern Slavonic and Sorbian allow perfective forms is a later innovation of these languages (Whaley 2000: 94-95).

‘material’ (Common Slavonic **bqdq* ‘to be’), which due to its semantics, ‘naturally’ evolved into a FUTURE TENSE marker in the (mostly Northern) Slavonic area, but with significantly different outputs.

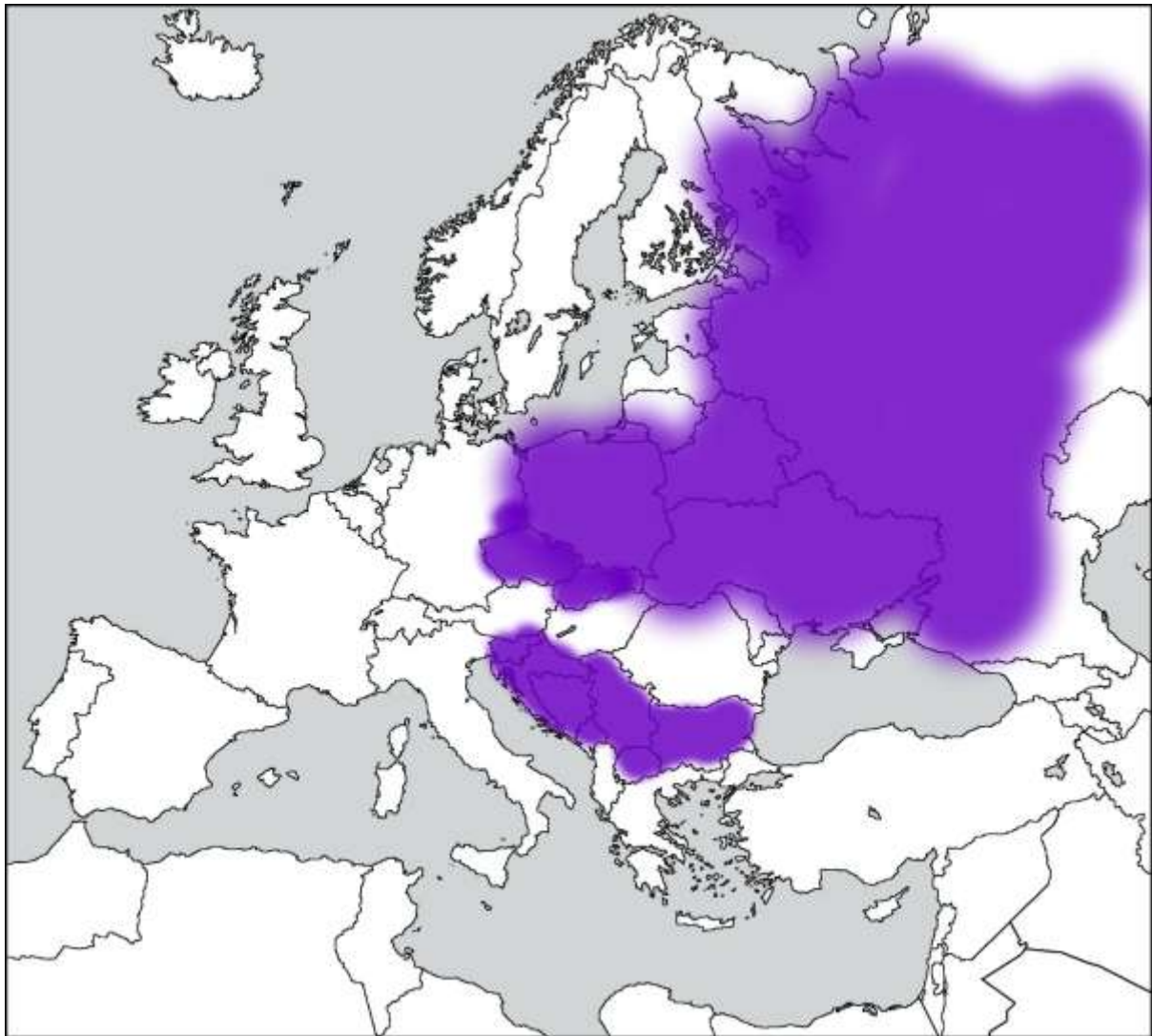


Figure 17 Distribution of COPULAR FUTURE. Sources (Dahl 2000b, Whaley 2000).

7.4.7. Hapax constructions

Among the constructions that have been documented in isolated utterances, the only construction worth mentioning is the verb **pustiti*[PST] ‘to let’ + INFINITIVE. Its low

frequency and **TAM** restrictions already make it very unlikely to be a grammaticalised FUTURE TENSE construction. Moreover, we can find further evidence for disregarding its legitimacy as a FUTURE TENSE construction by taking a quick look at its morpho-semantic structure and typology. From a typological point of view, the verb *pustiti* ‘to let’ (in its primary meaning) does not conform to an expected semantic base for a FUTURE construction (Bybee 1985, Ultan 1978), unlike the other FUTURE constructions analysed in this work.

On the one hand, the verb **pustiti* seems a good candidate to develop into an optative/desiderative (and perhaps even an imperative) auxiliary. For example, English *let*, particularly in archaic texts, developed into a marker of optativity, as in Shakespeare’s “*Let me not to the marriage of true minds*”. On the other hand, it is true that in the contexts where it has been documented (see (167) b-d.) it may express some intention, but it is hard to read this auxiliary as an optative. Hence, this points out the need for deeper research on this particular construction; due to my lack of sufficient data I cannot provide an answer now.

7.5. Summary

We have seen that there is morphological, semantic and typological evidence to state that the main six forms analysed here are genuine FUTURE TENSE REFERENCE constructions. This means that the FUTURE TENSE is quite grammaticalised in West Polesian (varying in degree from one construction to another), that is to say it is overtly marked (Dahl 2000b), and that all these six FUTURE TENSE constructions are

highly inflectional (less so the PERFECTIVE FUTURE). As with suppletive stems (Chapter 6), it can be said that West Polesian concentrates *all* the FUTURE TENSE constructions (semantic bases) present in all the Slavonic languages, except for the Slovenian FUTURE construction *bo-* + *l*-participle (Dahl 2000b: 324).³⁰⁴ Likewise, West Polesian integrates a significant number of the main FUTURE TENSE constructions (again, from a semantic perspective) present in European languages.

On the one hand, when discussing the suppletive forms of ‘year’ and ‘person’ (Chapter 6) we have seen that there is a three-way distinction, but that it is most likely motivated by language contact and socially-neutral variation. Now with the FUTURE TENSE, it is a six-way distinction (each form having more distinct functions than with the suppletive stems). Furthermore, the contact with the language groups from which the forms may have been borrowed is more distant in time. So how can we justify all this diversity of constructions?

On the other hand, according to Bybee & Dahl (1989) having several FUTURE TENSE constructions is a normal phenomenon:

“[...] at any one stage a language may have constructions that are close to one another semantically; e.g. it is not uncommon for a language to have more than one gram expressing the notions associated with future[...]” (Bybee & Dahl 1989: 96).

³⁰⁴ And again, according to Whaley (2000) this Slovenian *bo-* + INFINITIVE is one of the outputs or resulting variants of the copular form.

We know that in Common Slavonic the FUTURE TENSE was not as developed (i.e. grammaticalised) as in contemporary Slavonic languages. In fact, according to Dahl (2000c, 2000d) and Danylenko (2011) neither do some Indo-European languages have fully grammaticalised FUTURE TENSE, nor did Old Church Slavonic. Instead, in older stages of Slavonic multiple periphrastic constructions were employed, primarily with a de-obligative, inchoative and de-volitive meaning. We can see that Southern Slavonic and West Polesian have retained some vestiges of those periphrases, which have fully grammaticalised in contemporary Southern Slavonic, and at least partially (given that some retain modal values) in West Polesian. Yet there is also evidence that even in older stages of Russian more future constructions were used (which according to Whaley (2000: 122) retained their polysemy), but which have almost faded with time. There is also evidence that this has been the case throughout the history of Belarusian and Ukrainian up to our time (Avanesaŭ et al. 1963, Danylenko 2011, Jankoŭski 1989, Shevelov 1963).³⁰⁵ Aside from the abundance of FUTURE TENSE constructions, West Polesian is interesting for having preserved this ‘older stage’ of the language, whilst also having innovated in the use of these constructions.

Particularly, two ‘deviations’ should attract our attention. The first is the development of the REFLEXIVE DE-OBLIGATIVE and the SYNTHETIC FUTURE TENSE (also present in Ukrainian and Southwestern Belarusian), which have no precedents in the Slavonic family. The second is the unexpected correspondence (or divergent development) of the DE-VOLITIVE and DE-OBLIGATIVE constructions, which have switched the functions we would expect from their etymology and primary meaning.

³⁰⁵ Particularly in Southwestern Belarusian dialects.

Finally, we have mentioned the rare constructions with **pustiti* + infinitive. At first glance, we would expect it to be an auxiliary in the process of developing into a modal or a marker of optativity. However, within the few utterances documented in the corpus, it is far closer to futurity, which looks like an innovation deserving further study. However, I acknowledge the limitations of my corpus.

In sum, the West Polesian FUTURE TENSE is a cornucopia of still (somewhat) mysterious constructions, which I hope will attract the attention of more researchers in the coming years, while there are speakers left.

Part III. Summary and conclusions

Chapter 8

Summary and conclusions

I started by introducing the general situation of Western Polesie and West Polesian, whilst highlighting the literature gap (Chapter 1). I explained how I conducted this research project as well as the biases and the difficulties faced due to the interference of closely related varieties (Chapter 2). Then, I gave a bird's-eye overview of West Polesian phonology, morphology and syntax, stressing where it differs from other members of the Eastern Slavonic family (Chapter 3). Afterwards, I introduced the phenomenon of numerals and numeral phrases in West Polesian (Chapter 4). The next chapter (Chapter 5) has focused on the odd government and agreement triggered by cardinal numerals (particularly lower numerals) and the peculiar morphosyntactic nature of the adnumerative forms. I continued with suppletion (mainly) motivated by quantification in the nouns 'year' and 'person', and I compared them in the light of the canonical instance (Chapter 6). In the next chapter (Chapter 7), I presented several future tense reference constructions I documented in West Polesian and I debated their legitimacy as 'inflectional forms' based on grammaticality tests and a cross-linguistic and historical comparison of these forms. Finally, in this chapter (Chapter 8) I present a summary of the results and tie together loose threads from the conclusions extracted in the previous chapters.

The main contributions of this thesis can be divided into three areas:

1. Documenting and describing the morphology and syntax of West Polesian
2. Dealing with the linguistic interferences caused by several closely related languages (in the field).
3. Analysing theoretical questions that arise in the light of data from West Polesian.

First, I have documented what is special about West Polesian morphology and syntax, in comparison with the Eastern Slavonic family and cross-linguistically. I summarise the main morphological and/or syntactic peculiarities of West Polesian in the following list:

- Postnominal possessors with HUMAN nouns (§3.3.).
- ANIMACY split in the PLURAL with nouns denoting animals, or a differentiation between INANIMATES - ANIMATES (where farm animals usually belong) - HUMANS (§3.4.; §4.3.2.3.; Huntley (1980)).
- Pronominal numerals (§4.4.).
- An ADNUMERATIVE form available for every countable noun followed by a lower numeral (§5.2.2.).
- Multiple choices for adjective agreement (with lower numerals and ADNUMERATIVES) not exclusively available for FEMININE nouns (§5.2.3.1.1.).
- A type of SECOND GENITIVE used with quantification and higher numerals available for some nouns (§5.3.).

- Up to three suppletive stems existing for the nouns ‘year’ and ‘person’ (Chapter 6).
- Six FUTURE-TIME-REFERENCE constructions (Chapter 7).
- A reflexive form of the DE-OBLIGATIVE FUTURE (§7.4.2.).
- A semantic shift between the DE-OBLIGATIVE and DE-VOLITIVE FUTURE tense constructions (§7.4.3.).

Second, in this project I have been constantly walking on a tightrope shaken by cross-linguistic interferences, as I had to judge which data were genuine West Polesian forms. This generated lots of noise, hence, besides participatory observation and overheard conversations, I often used the criticisms of the recordings made by other speakers in the village as a guide. Yet, on the one hand, West Polesian not being standardised is more open to innovations and variations than its standardised neighbours. On the other hand, being in an extended diglossic situation in which the prestigious varieties are closely related (in a continuum) to the basilect or less prestigious variety, there is an easy way to borrow constructions and forms from the prestigious varieties.

Related to this, I also had to deal with sociologically-neutral inter- and intra-speaker variation, which in part is also responsible for the multiple morphosyntactic agreement possibilities, overabundant inflectional paradigms and rich future tense construction inventories. As already pointed out in (§6.6.), Dorian (2010: 312) says that sociolinguistic situations in which there is sociologically-neutral variation are far

more common than linguists tend to imagine, and Western Polesie is just an example of this (with the addition that all the surrounding varieties are genetically related). Moreover, overabundance has been present in many parts of this work. This may be hard to accept for linguists who believe that absolute synonyms do not exist and who seek for additional (hidden) meanings and functions in order to justify the existence of the additional forms. Thus, I reiterate the call to embrace variation and accept that overabundance exists in natural languages.

Third, the main input from typology overarching the main topics of this research is setting a canon or a scale in order to compare phenomena (within the system itself as well as cross-linguistically). I took a canonical approach (CT) for phenomena related to quantification (canonical CASE, NUMBER and FEATURE values; and canonical suppletion). Beyond CT, I have also used grammaticalisation tests for FUTURE TENSE REFERENCE constructions, eventually also trying to determine how genuine or canonical they are as inflectional constructions. This has opened the door to a broader theoretical question, which is how to describe phenomena that are in the penumbra, particularly the ADNUMERATIVE; a non-binary value. Nevertheless, the expectation from CT was already that it is unlikely that a parameter in a given language will meet all the canonicity criteria (e.g. there will be no language with a canonically suppletive verb or noun). The results from the analyses, which show all the parameters on a spectrum rather than an absolute endpoint, are not surprising. In fact, according to Bybee's (1985) and Ultan's (1978) research, this is the expectation for future tense constructions cross-linguistically (i.e. to have multiple constructions in different levels of grammaticalisation).

In sum, the data and analyses in this work only arrive up to a certain point. Yet my hope is they will serve as a platform from which future research projects on West Polesian can be launched.

References

- ACQUAVIVA, Paolo, 2006. Goidelic inherent plurals and the morphosemantics of number. *Lingua* vol. 116, no. 11, pp. 1860-1887. DOI 10.1016/j.lingua.2004.10.003.
- ACQUAVIVA, Paolo, 2008. *Lexical plurals*. Oxford: Oxford University Press.
- AKINER, Shirin, 1983. The Syntax of the Numeral in Byelorussian, Compared with Ukrainian, Russian and Polish. *The Slavonic and East European Review* vol. 61, no. 1, pp. 55-68.
- ALEXANDER, Ronelle, 2006. *Bosnian, Croatian, Serbian, a Grammar: With Sociolinguistic Commentary*. Madison, WI: University of Wisconsin Press.
- ARONOFF, Mark, 1994. *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, MA/London: MIT Press.
- ARONOFF, Mark, 2016. Competition and the lexicon. In: ELIA, A., IACOBINI, C. and VOGHERA, M., eds. *Livelli di Analisi e fenomeni di interfaccia. XLVII Congresso Internazionale della Società di Linguistica Italiana* 39-52.
- ARWYN WATKINS, T., 1993. Welsh. In: BALL, M. J. and J. FIFE eds., *The Celtic languages*. London: Routledge, pp. 289-348.
- AVANESAŮ, R. I., 1964. *Narysy pa belaruskaj dyjalektalohii*. Minsk: Navuka i tèxnika.
- AVANESAŮ, R. I., KRAPIVA, K. K. and MACKEVIČ, Ju F., 1963. *[DABM] Dyjalektalahičny atlas belaruskaj movy (vol. I-II)*. Minsk: AN BSSR.
- BABBY, Leonard H., 1987. Case, Prequantifiers, and Discontinuous Agreement in Russian. *Natural Language & Linguistic Theory* vol. 5, no. 1, pp. 91-138. DOI 10.1007/BF00161869.
- BAERMAN, Matthew, BROWN, Dunstan and CORBETT, Greville G., 2005. *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge: Cambridge University Press.

- BAILYN, John and NEVINS, Andrew I., 2008. Russian Genitive Plurals are Impostors. In: BACHRACH, A. and NEVINS, A. eds., *Infectional identity*. Oxford: Oxford University Press, pp. 237-270.
- BARSZCZEWSKA, Nina, 2016. Zachowanie językowe mieszkańców Podlasia w warunkach bilingwizmu: Charakterystyka socjolingwistyczna. *Acta Albaruthenica* vol. 16, pp. 143-168.
- BEVZENKO, Stepan F., 1980. *Ukrajinska dialektolohija*. Kiev: Vyšča škola.
- BIRYLA, Mikalaj V., 1986. *Nacisk nazoŭnikaŭ u sučasnaj belaruskaj move*. Minsk: Vyšějšaja Škola.
- BOBALJIK, Jonathan D., 2012. *Universals in comparative morphology: suppletion, superlatives, and the structure of words*. Cambridge, MA: MIT Press.
- BOND, Oliver, 2018. Canonical Typology. In: AUDRING, J. and MASINI, F. eds., *The Oxford handbook of morphological theory*. Oxford: Oxford University Press, pp. 409-431.
- BORSLEY, Robert D., TALLERMAN, Maggie and WILLIS, David, 2007. *The Syntax of Welsh*. Cambridge: Cambridge University Press.
- BORTNIK, Ninel', 1979. Nekotorye osobennosti osnov obrazujuščix dvojnye formy množestvennogo čisla. *Russian Language Journal* vol. 32, pp. 43-58.
- BOSAK, A. and BOSAK, V., 2005. *Atlas havorak Pružanskaha rajona Brėtskaj voblastsi i sumežža (Verxnjaha Nadjasel'dzja)*. *Fanetika i marfalohija*. Minsk: Nacyjanal'na Akadėmija Navuk Belarusi.
- BROWN, Dunstan, 1998. Defining 'subgender': Virile and devirilized nouns in Polish. *Lingua* vol. 104, no. 3-4, pp. 187-233.
- BROWN, Dunstan, 2007. Peripheral Functions and Overdifferentiation: The Russian Second Locative. *Russian Linguistics* vol. 31, no. 1, pp. 61-76. DOI 10.1007/s11185-006-0715-5.

- BROWN, Dunstan and CHUMAKINA, Marina, 2012. What there is and what there might be: An introduction to Canonical Typology. In: BROWN, D., CHUMAKINA, M. and CORBETT, G. G. eds., *Canonical morphology and syntax*. Oxford: Oxford University Press, pp. 1-19.
- BUSLAEV, F. I., 1959. *Istoričeskaja grammatika russkogo jazyka*. 2nd ed. Moscow: Gosudarstvennoe učebno-pedagogičeskoe izd-vo.
- BYBEE, Joan L., 1985. *Morphology. A Study of the Relation between Meaning and Form*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- BYBEE, Joan L. and DAHL, Östen, 1989. The Creation of Tense and Aspect Systems in the Languages of the World. *Studies in Language* vol. 13, no. 1, pp. 51-103. DOI 10.1075/sl.13.1.03byb.
- CAPPELLARO, Chiara, 2013. Overabundance in diachrony: A case study. In: CRUSCHINA, S., MAIDEN, M. and SMITH, J. C. eds., *The boundaries of pure morphology: Diachronic and synchronic perspectives*. Oxford: Oxford University Press, pp. 209-220.
- CHUMAKINA, Marina, HIPPISEY, Andrew and CORBETT, Greville G., 2004. Istoričeskie izmenenija v rusškoj leksike: slučaj čeredujuščegosja suppletivizma. *Russian Linguistics*, no. 28, pp. 281 – 315.
- COMRIE, Bernard, 2013. Numeral Bases. In: Dryer, Matthew S. & Haspelmath, Martin ed., *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- COMRIE, Bernard and CORBETT, Greville G., 1993. *The Slavonic languages*. London/New York: Routledge.
- CORBETT, Greville G., 1978a. Apposition involving dva, tri, četyre in Russian - A solution to Worth's Riddle. *Quinquereme -New Studies in Modern Languages* vol. 1, no. 2, pp. 258-264.
- CORBETT, Greville G., 1978b. Numerous Squishes and Squishy Numerals in Slavonic. In: COMRIE, B. ed., *Classification of grammatical categories*. Edmonton, Canada: Linguistic research INC., pp. 43-73.

- CORBETT, Greville G., 1978c. Problems in the Syntax of Slavonic Numerals. *The Slavonic and East European Review* vol. 56, no. 1, pp. 1-12.
- CORBETT, Greville G., 1978d. Universals in the syntax of cardinal numerals. *Lingua* vol. 46, no. 4, pp. 355-368. DOI 10.1016/0024-3841(78)90042-6.
- CORBETT, Greville G., 1983. *Hierarchies, targets and controllers: agreement patterns in Slavic*. London: Croom Helm.
- CORBETT, Greville G., 1991. *Gender*. Cambridge: Cambridge University Press.
- CORBETT, Greville G., 1993. The head of Russian numeral expressions. In: CORBETT, G. G., FRASER, N.M. and MCGLASHAN, S. eds., *Heads in grammatical theory*. Cambridge: Cambridge University Press, pp. 11-35.
- CORBETT, Greville G., 1996. Minor number and the plurality split. *Rivista Di Linguistica* vol. 8, pp. 101–122.
- CORBETT, Greville G., 2000. *Number*. Cambridge: Cambridge University Press.
- CORBETT, Greville G., 2001. Number. In: HASPELMATH, M. ed., *Language typology and language universals: An international handbook*. Berlin: Walter de Gruyter, pp. 816-830.
- CORBETT, Greville G., 2005. The canonical approach to typology. In: Zygmunt Frajzyngier, Adam Hodges and David S. Rood eds., *Linguistic diversity and language theories*. Amsterdam: John Benjamins, pp. 25-49.
- CORBETT, Greville G., 2007. Canonical typology, suppletion, and possible words. *Language* vol. 83, no. 1, pp. 8-42.
- CORBETT, Greville G., 2008. Determining morphosyntactic feature values: the case of case. In: CORBETT, G. and M. NOONAN eds., *Case and grammatical relations: Papers in honour of Bernard Comrie*. Amsterdam: John Benjamins.
- CORBETT, Greville G., 2010. Canonical derivational morphology. *Word Structure* vol. 3, pp. 141–55. DOI 10.3366/E1750124510000565.

- CORBETT, Greville G., 2011. The penumbra of morphosyntactic feature systems. *Morphology* vol. 21, no. 2, pp. 445–480. DOI 10.1007/s11525-010-9171-4.
- CORBETT, Greville G., 2012. *Features*. Cambridge: Cambridge University Press.
- CORBETT, Greville G., 2013. Canonical morphosyntactic features. In: BROWN, D., M. CHUMAKINA and G. CORBETT eds., *Canonical Morphology & Syntax*. Oxford: Oxford University Press, pp. 48-65.
- CUBBERLEY, Paul, 1993. Alphabets and transliteration. In: COMRIE, B. and G.G. CORBETT eds., *The Slavonic Languages*. London/New York: Routledge, pp. 20-59.
- DAHL, Östen, 2000a. The Future Time Reference Questionnaire. In: DAHL, Ö ed., *Tense and Aspect in the Languages of Europe*. Berlin/New York: Mouton de Gruyter, pp. 789-799.
- DAHL, Östen, 2000b. The grammar of future time reference in European languages. In: DAHL, Ö ed., *Tense and Aspect in the Languages of Europe*. Berlin/New York: Mouton de Gruyter, pp. 309-327.
- DAHL, Östen, 2000c. The tense-aspect systems of European languages in a typological perspective. In: DAHL, Ö ed., *Tense and Aspect in the Languages of Europe*. Berlin/New York: Mouton de Gruyter, pp. 3-25.
- DAHL, Östen, 2000d. Verbs of becoming as future copulas. In: DAHL, Ö ed., *Tense and Aspect in the Languages of Europe*. 2000 2000 ed. Berlin/New York: Mouton de Gruyter, pp. 351-361.
- DANYLENKO, Andrii, 2011. Is there any inflectional future in East Slavic? A case of Ukrainian against Romance reopened. In: NOMACHI, M. ed., *Grammaticalization in Slavic Languages: From Areal and Typological Perspectives*. Sapporo: Slavic Research Center, Hokkaido University, pp. 147-177.
- DEL GAUDIO, Salvatore, 2014. Per un approccio preliminare ai dialetti di transizione ucraino-bielorussi: l'area nord-occidentale della regione di Černihiv. In: BONOLA, A., RAMUSINO, P. C. and GOLETIANI, L. eds., *Studi italiani di linguistica slava. Strutture, uso e acquisizione*. Florence: Firenze University Press, pp. 273-288.

- DINGWALL, William O., 1969. Government, Concord and Feature-Change Rules. *Glossa* vol. 3, no. 2, pp. 210-240.
- DIXON, Robert M. W., 1982. *Where have all the adjectives gone? and other essays in semantics and syntax*. Berlin/New York: Mouton.
- DOBROVSKÝ, Josef, 1822. *Institutiones linguae slavicae dialecti veteris quae quum apud russos, serbos aliosque ritus graeci, tum apud dalmatas glagolitas ritus latini slavos in libris sacris obtinet. Cum tabulis aeri incisis quatuor*. Vienna.
- DORIAN, Nancy C., 2010. *Investigating Variation. The Effects of Social Organization and Social Setting*. New York/Oxford: Oxford University Press.
- DRYER, Matthew and HASPELMATH, Martin, 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- DUCHET, Jean-Louis, 1995. The Albanian Tense System. In: THIEROFF, R. ed., *Tense Systems in European Languages II*. Berlin/Boston: De Gruyter, Inc., pp. 253-276.
- DULIČENKO, Aleksandr D., 1981. *Slavjanske literaturnye mikrojazyki. Voprosy formirovanija i razvitija*. Tallin: Valgus.
- DULIČENKO, Aleksandr D., 1995. The West Polesian Literary Language. *Uppsala Multiethnic Papers* vol. 34, pp. 119-131.
- ECKHOFF, Hanne, 2015. Animacy and differential object marking in Old Church Slavonic. *Russian Linguistics; International Journal for the Study of Russian and Other Slavic Languages* vol. 39, no. 2, pp. 233-254. DOI 10.1007/s11185-015-9148-3.
- ENGELKING, Anna, 2001. Poleska ekspedycja etnosocjologiczna Józefa Obrębskiego w latach 1934-1937. Organizacja. Metody badań. Problematyka. Uczestnicy. *Etnografia Polska* vol. 45, no. 1-2, pp. 23-45.
- ENGELKING, Anna, 2006. Po swojemu, po polsku, po białorusku. Praktyka językowa a poczucie tożsamości narodowej prawosławnych mieszkańców pogranicza polsko-białoruskiego. In: KURCZEWSKA, J. ed., *Oblicza lokalności. Różnorodność miejsc i czasu*. Warsaw: IFiS PAN, pp. 26-46.

- EVANS, Nicholas D., 1995. *A grammar of Kayardild: With historical-comparative notes on Tangkic*. Berlin/New York: Mouton de Gruyter.
- FRANKS, Steven, 1995. *Parameters of Slavic morphosyntax*. Oxford University Press.
- GOLACHOWSKA, Ewa, 2006. *Język i kultura mieszkańców wsi wsłociańskich i szlacheckich dawnej ziemi Drohickej i na Podlasiu (Studium Socjolingwistyczne)*. Warsaw: Sławystyczny Ośrodek Wydawniczy.
- GOOD, Jeff, 2011. Data and Language Documentation. In: AUSTIN, P. and J. SALLABANK eds., *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press, pp. 212-234.
- GRACE, George W., 1992. How Do Languages Change? (More on "Aberrant" Languages). *Oceanic Linguistics* vol. 31, no. 1, pp. 115-130. DOI 10.2307/3622968.
- GREENBERG, Joseph H., 1978. Generalizations about Numeral Systems. In: GREENBERG, J. H., Ferguson, A. CHARLES and E.A. MORAVCSIK eds., *Universals of human language*. Stanford UP, pp. 249-295.
- GUSTAVSSON, Sven, 1998. Sociolinguistic Typology of Slavic Minority Languages. *Slovo: Journal of Slavic Languages and Literatures*, no. 46, pp. 75-89.
- GVOZDANOVIĆ, Jadranka, 1994. The Tense System of Russian. In: THIEROFF, R. and J. BALLWEG eds., *Tense Systems in European Languages*. Tübingen: Niemeyer, pp. 191-200.
- GVOZDANOVIĆ, Jadranka, 1995. Western South Slavic Tenses in a Typological Perspective. In: THIEROFF, R. ed., *Tense Systems in European Languages II*. Berlin/Boston: De Gruyter, Inc, pp. 181-194.
- HEDIN, Eva, 1995. The Tense-Aspect System of Modern Greek. In: THIEROFF, R. ed., *Tense Systems in European Languages II*. Berlin/Boston: De Gruyter. Inc., pp. 233-252.
- HULICKAJA, V., ZLOBIN, L. and RUDAKOŬSKAJA, Z., 1992. *Belaruskaja i ruskaja dyjalektalohija*. Minsk: Vyšėjšaja Škola.

- HUNTLEY, David, 1980. *The Evolution of Genitive-Accusative Animate and Personal Pronouns in Slavic Dialects*. The Hague: Mouton.
- HURSKI, M., 1972. *Paraŭnal'naja hramatyka ruskaj i belaruskaj moŭ*. 2nd ed. Minsk: Vyšėjšaja škola.
- HYMAN, Larry M., 2012. In defense of Prosodic Typology: A Response to Beckman and Venditti. *Linguistic Typology* vol. 16, no. 3, pp. 341-385.
- IGARTUA, Iván and MADARIAGA, Nerea, 2018. The interplay of semantic and formal factors in Russian morphosyntax: animate paucal constructions in direct object function. *Russian Linguistics* vol. 42, no. 1, pp. 27-55. DOI 10.1007/s11185-017-9188-y.
- JAKOBSON, Roman, 1932. *Zur Struktur der russischen Verbums*. Prague: Pražský Lingvistický Kroužek.
- JANKOŮSKI, F. M., 1980. *Sučasnaja belaruskaja litaraturnaja mova. Marfalohija*. 2nd ed. Minsk: Vyšėjšaja škola.
- JANKOŮSKI, F. M., 1989. *Histaryčnaja hramatyka belaruskaj movy*. 3rd ed. Minsk: Vyšėjšaja škola.
- JANKOWIAK, Mirosław, 2014. Belarusian Dialects in Latvian Latgale—Transitional or Mixed. In: SERŽANT, I. A. and WIEMER, B. eds., *Contemporary Approaches to Dialectology. The Area of North, North-West Russian and Belarusian Dialects*. Bergen: Slavica Bergensia, pp. 344-372.
- JANKOWIAK, Mirosław and GRAJEWSKI, Łukasz, 2011. Szlakiem białoruskich Poleszuków (raport z badań terenowych). *Acta Baltico-Slavica* vol. 35, pp. 215-236. DOI 10.11649/abs.2011.016.
- KASCJUČYK, Valjancina M., 2005. Varyatnasc' u dyjalektnaj frazealohii (Na asнове havorak belarуска-ukrainskaha pamežža). In: ALJAKHOVIČ, M., ed. *Matèryaly IV Mižnarodaj Navukovaj Kanferèncij. Movy Vjalikaha Knjastva Litoŭskaha*, pp. 182-184.

- KENESEI, István, VAGO, Robert M. and FENYVESI, Anna, 1998. *Hungarian*. London: Routledge.
- KIBORT, A., 2008. *A typology of grammatical features*. Available from: <http://www.grammaticalfeatures.net/inventory.html>.
- KIM, Hyounsup, 2009. *The structure and use of collective numeral phrases in Slavic: Russian, Bosnian/Croatian/Serbian, and Polish* [PhD Thesis]. The University of Texas at Austin.
- KLIMČUK, Fëdar D., 1973. Nekatoryja paleska-paŭdnëvaslavjanskija paraleli. In: *Typalohija i ŭzaemadzejanne slavjanskix moŭ i litaratur. II-R ësrepublikańskaja kanfer èncyja*, pp. 129-130.
- KLIMČUK, Fëdar D., 1983. *Havoraki Zaxodnjaga Palessja: Fanetyčny narys 4-ox havorak*. Minsk: Navuka i t èxnika.
- KLIMČUK, Fëdar D., 1992. Heradot pra Palesse. In: *Prablemy narodnaj adukacyj, navuki, kul'tury belaruskaha Palessja*, pp. 79-81.
- KLIMČUK, Fëdar D., 1999. Poleskij toponim Zagorod'e. In: *Zaharoddze: Materyaly Miždyscyplinarna navukovaha seminaru pa pytanjax dasledavannja Palessja*, pp. 10-23.
- KLIMČUK, Fëdar D., 2006. Slavjanskija i indaeŭrapejskija suvjazy tarakanskix havorak (r èfleks *e, *ъ > [a]). *Zora* vol. 41, no. Diahronija in sinhronija v dialektoloških raziskavah, pp. 296-302.
- KORJAKOV, Jurij B., 2002. *Jazykovaja situacija v Belorussii i tipologia jazykovyx situacii*. Moscow State University.
- KOZHANOV, Kirill A., 2016. Analitičeskoe buduščee vremja v jazyke russkix cygan kak kal'ka s vostočno-slavjanskix jazykov. *Slovene* vol. 5, no. 1, pp. 249-262.
- KUTEVA, Tania, 1995. Bulgarian Tenses. In: THIEROFF, R. ed., *Tense Systems in European Languages II*. Berlin/Boston: De Gruyter, Inc, pp. 195-213.

- LEISS, Elisabeth, 1985. Zur Entstehung des neuhochdeutschen analytischen Futurs. *Sprachwissenschaft* vol. 10 (3-4), pp. 250-73.
- LEKO, Nedžad, 2009. *The syntax of numerals in Bosnian*. Munich: Lincom Europa.
- LEPEŠAŬ, I., 2002. *Sučasnaja belaruskaja litaraturnaja mova: sprèčnyja pytanni*. Hrodna: HrDU im. Jan Kupala.
- LEVANCÈVIČ, Lena V., 1994. *Atlas havorak Bjarozaŭskaha raëna Brèstskaj vovlastsi. Marfalahija*. Brèski Dzjaržaŭny Pedahahičny Instytut im, A. S. Puškina.
- LEVANCÈVIČ, Lena V., 2013. *Linhvističnae krajasnaŭstva Brèstčyny*. Brest: BrDU A.S. Puškina.
- LJAPĚŠKIN, V. U., 1989. Nekatoryja dubljetnyja sklonavyja formy nazoŭnikaŭ belaruskaj i ukrajnskaj move. *Belaruskaja Linhvistyka* vol. 35, pp. 44-50.
- LJAPĚŠKIN, V. U., 1990. Kul'tura movy i svabodnae var'iravanne kančatkaŭ nazoŭnikaŭ. *Vesci AN BSSR. Serya Hramadskix Navuk* vol. 5, pp. 107-113.
- LJAPĚŠKIN, V. U., 1991a. Kadyfikacyja i status nekatoryx hramatyčnyx formaŭ. *Belaruskaja Linhvistyka* vol. 39, pp. 68-70.
- LJAPĚŠKIN, V. U., 1991b. Varyjantnasc' form nazoŭnikaŭ rodnaha sklonu množnaha liku. *Pytanne kul'tury pis'movaj movy*, pp. 137-146.
- LOMTEV, T. P., 1956. *Grammatika belorusskogo jazyka. Posobie dlja universitetov i pedagogičeskix institutov*. Moscow: Gos. Učebno-Pedagogičeskoe izdatel'stvo Ministerstva Prosveščeniya RSFSR.
- MACKEVIČ, Ju F., 1959. *Marfalahija dzejaslova ŭ belaruskaj move*. Minsk: AN BSSR.
- MACKEVIČ, Ju F., MURAŠKO, A. and OREŠKOVA, A., 1964. Ob itogax raboty po sostavleniju atlasa belorusskix narodnix govorov. In: AVANESOV, R. I. ed., *Voprosy dialjektologii vostočno-slavjanskikh jazykov*. Minsk: Navuka i tèxnika, pp. 39-59.

- MACRI, Martha and SARMENTO, James, 2010. Respecting Privacy: Ethical and Pragmatic Considerations. *Language & Communication* vol. 30, no. 3, pp. 192-197. DOI 10.1016/j.langcom.2009.11.005.
- MADARIAGA, Nerea and IGARTUA, Iván, 2017. Idiosyncratic (Dis)agreement Patterns: The Structure and Diachrony of Russian Paucal Subjects. *Scando-Slavica* vol. 63, no. 2, pp. 99-132. DOI 10.1080/00806765.2017.1390922.
- MALAŽAI, H. and KLIMČUK, F., 1989. *Dyjalektny sloŭnik Brèstčyny*. Minsk: Navuka i Tèxnika.
- MATTHEWS, P. H., 1991. *Morphology*. 2nd ed. Cambridge: Cambridge University Press.
- MAYER, Gerald, 1971. Ukrainian Dva (Dvi), Try, Čotyry + Noun: a Reinterpretation. *Canadian Slavonic Papers* vol. 13, pp. 73-79.
- MAYO, Peter J., 1976. *A grammar of Byelorussian*. Sheffield: Anglo-Byelorussian Society, in association with the Dept. of Russian and Slavonic Studies, University of Sheffield.
- MAZZITELLI, Lidia F., 2011. Possession, modality and beyond: The case of mec' and mecca in Belarusian. In: NOMACHI, M. ed., *Grammaticalization in Slavic Languages: From Areal and Typological Perspectives*. Sapporo: Slavic Research Center, Hokkaido University, pp. 179-201.
- MEILLET, Antoine, 1912. L'évolution des formes grammaticales. *Scientia* 6(12), pp. 384-400.
- MEL'ČUK, Igor A., 1985. *Poverxnostnyj sintaksis russkix čislovyx vyraženij*. Vienna: Institut für Slawistik der Universität Wien.
- MEL'ČUK, Igor A., 1994. Suppletion: toward a logical analysis of the concept. *Studies in Language* vol. 18, no. 2, pp. 339-410.
- MIECHOWICZ-MATHIASSEN, Katarzyna, 2011. *The syntax of Polish cardinal numerals* [MA Thesis]. Adam Mickiewicz University.

- MIKHAÏLYK, R., 2003. *Grammaire pratique de l'ukrainien: Manuel du niveau moyen*. Paris: L'Harmattan.
- MILROY, James, 1999. The consequences of standardization in descriptive linguistics. In: BEX, T. and R.J. WATTS eds., *Standard English: The widening debate*. London: Routledge, pp. 16-35.
- MÖNKE, H., 1971. *Das Futurum der polnischen Verba*. Munich: Otto Sagner.
- MOSZYŃSKI, Leszek, 2006. *Wstęp do filologii słowiańskiej*. 2nd ed. Warsaw: Wydawnictwo Naukowe PWN.
- NARKEVIČ, A., 1976. *Nazoŭnik. Hramatyčnyja katèhoryi i formy*. Minsk: BDU im. Lenina.
- NURMIO, Silva and WILLIS, David, 2016. The rise and fall of a minor category: The case of the Welsh numerative. *Journal of Historical Linguistics* vol. 6, no. 2, pp. 297–339. DOI 10.1075/jhl.6.2.05nur.
- OSTHOFF, Hermann, 1888. Etymologica I. *Beiträge Zur Geschichte Der Deutschen Sprache Und Literatur* vol. 13, pp. 395-463.
- OSTHOFF, Hermann, 1899. *Vom Suppletivwesen der indogermanischen Sprache*. Heidelberg: Universitätsbuchdruckerei von J. Hörning.
- PALJAKOŭ, A., 2013. Zaxadnapaleskaja mova. In: CYXUN, H. A. ed., *Zamežnaja movaznaŭčaja belarusistyka na mižnarodnyx z'jezdax slavistaŭ. Zbornik dakladaŭ da XV Mižnarodnaga Z'ezda Slavistaŭ*. Minsk: Belaruskaja navuka, pp. 122-133.
- PAULUS, J., 1970. Change of e > 'o in the dialect of Northwest Polissye, BSSR. In: *Actes du Xe Congrès International des Linguistes* (offprint). Bucarest, pp. 221-225.
- PERELTSVAIG, Asya, 2009. Babby's puzzle: syntax or morphology? In: *Formal Approaches to Slavic Linguistics - 18th Meeting*.
- PERELTSVAIG, Asya, 2013. On Number and Numberlessness in Languages without Articles. In: CATHCART, C., CHENG, I., FINLEY, G., KANG, S., SANDY, C. S. and

- STICKLES, E., eds. *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society*, pp. 300–314.
- PESETSKY, David, 2013. *Russian Case Morphology and the Syntactic Categories*. Cambridge, MA; London: The MIT Press.
- PISKUNOŮ, Fëdar, 2010. Dva harady ci dva horady. Ličëbniki dva, try, čatyry ŭ kol'kasna-imennyyx spalučënnjax. *Rodnae Slova*, no. 6, pp. 51-53.
- PUGH, Stefan, 2009. *The Rusyn language*. Munich: LINCOM GmbH.
- PUGH, Stefan and PRESS, Ian, 1999. *Ukrainian: a comprehensive grammar*. London: Routledge.
- RAPPAPORT, Gilbert C., 2002. Numeral Phrases in Russian: A Minimalist Approach. *Journal of Slavic Linguistics* vol. 10, no. 1/2, pp. 327-340.
- RAPPAPORT, Gilbert C., 2003. The Grammatical Role of Animacy in a Formal Model of Slavic Morphology. In: *American Contributions to the Thirteenth International Congress of Slavists* pp. 149-166.
- RONCERO, Kristian, 2015. *West Polesian Inflectional Morphology: A preliminary approach* [MA Thesis]. University of the Basque Country.
- RONCERO, Kristian, 2016. Working paper on the vocative in West Polesian. In: GUSAKOV, V. G., ed. *Problemy racional'nogo ispol'zovaniya prirodnnykh resursov i ustoičivoie razvitie Poles'ja* vol II, pp. 672-676.
- ROUNDS, Carol, 2001. *Hungarian: an essential grammar*. London: Routledge.
- RUSANOVSKIJ, V. M., ŽOVTOBRJUX, M. A., GORODEMSKAJA, E. G. and GRIŠČENKO, A. A., 1986. *Ukrainskaia grammatika*. Kiev: Naukova dumka.
- SAKEL, Jeanette and EVERETT, Daniel, 2012. *Linguistic fieldwork: a student guide*. In: EVERETT, D. L. ed., Cambridge/New York: Cambridge University Press.
- SALMINEN, Tapani, 2010. Europe and the Caucasus. In: MOSELEY, C. ed., *Atlas of the World's languages in danger*. 3rd ed. Paris: UNESCO Publishing, pp. 32-42.

- SCATTON, Ernest A., 2002. Bulgarian. In: COMRIE, B. and G.G. CORBETT eds., *The Slavonic Languages*. 2nd ed. London/New York: Routledge, pp. 188-248.
- SCHUSTER-ŠEWC, H., 1999. *Grammar of the upper Sorbian language: phonology and morphology*. 2nd ed. München; Newcastle: Lincom Europa.
- SHEVELOV, George Y., 1964. *A prehistory of Slavic. Historical phonology of Common Slavic*. Heidelberg: Winter Univ. Verl.
- SHEVELOV, George Y., 2015. On the Chronology of *h* and the New *g* in Ukrainian. *Harvard Ukrainian Studies* vol. 1, no. 2, pp. 137-152. DOI 10.1163/2405-8262_rgg4_SIM_09355.
- SHEVELOV, George Y., 1963. *The Syntax of Modern Literary Ukrainian: The Simple Sentence*. The Hague: Mouton.
- SIMS-WILLIAMS, Nicholas, 1979. On the Plural and Dual in Sogdian. *Bulletin of the School of Oriental and African Studies, University of London* vol. 42, no. 2, pp. 337-346.
- SIMS-WILLIAMS, Nicholas, [Forthcoming in a Festschrift]. *Dual and Numerative in Middle and New Iranian*.
- SMITH-STARK, T., 1974. The Plurality Split. *Chicago Linguistics Society* vol. 10, no. 1, pp. 657-671.
- STUMP, Gregory T., 2006. Heterocclisis and paradigm linkage. *Language* vol. 82, no. 2, pp. 279-322.
- SUPRUN, Adam E., 1961. *Staroslavjanske čislitel'nye*. Frunze: Kirgizskij Gosudarstvennyj Universitet.
- SUSSEX, Roland and CUBBERLEY, Paul, 2006. *The Slavic Languages*. Cambridge: Cambridge University Press.
- TARAŠKEVIČ, B., 1918. *Belaruskaja hramatyka dlja škol*. Vilnius: Drukarnja M. Kuxty.

- THIEBERGER, Nick, 2012. *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press.
- THOMPSON, Sandra A., 1989. A discourse approach to the cross-linguistic category 'adjective'. In: CORRIGAN, R., F.R. ECKMAN and M. NOONAN eds., *Linguistic categorization*. Amsterdam: John Benjamins, pp. 245-266.
- THORNTON, Anna M., 2011. Overabundance (multiple forms realising the same cell): A non-canonical phenomenon in Italian verb morphology. In: MAIDEN, M., CHARLES SMITH, J., GOLDBACH, M. and HINZELIN, M. eds., *Morphological autonomy: Perspectives from Romance inflectional morphology*. Oxford: Oxford University Press, pp. 357-381.
- THORNTON, Anna M., 2013. La non canonicità del tipo It. braccio // braccia / bracci: sovrabbondanza, difettività o iperdifferenziazione? *Studi Di Grammatica Italiana* vol. 29-30, pp. 419-477.
- THORNTON, Anna M., 2019. Overabundance: a Canonical Typology. In: RAINER, F., GARDANI, F., LUSCHÜTZKY, H. and DRESSLER, W. U. eds., *Competition in inflection and word-formation*. Dordrecht: Springer, pp. 189-224.
- TOLSTOJ, N. I., 1968. *Leksika Poles'ja: materialy dlja poleskogo dialektного slovarja*. Moscow: Nauka.
- TOLSTOJ, N. I., 1983. Ètnogenetičeskie aspekti izučenie Poles'ja. In: TOLSTOJ, N. I., VINOGRADOVA, L. N. and TOLSTAJA, S. M. eds., *Poles'e i ètnogenez slavyan*. Moscow: Nauka, pp. 6-8.
- ULTAN, Russell, 1978. The Nature of Future Tenses. In: GREENBERG, J. H., FERGUSON, C. A. and MORAVCSIK, E. A. eds., *Universals of Human Language*. Stanford, CA: Stanford UP, pp. 83-123.
- VAILLANT, André, 1958. *Grammaire comparée des langues slaves (vol II): Morphologie. Deuxième partie, Flexion pronominale*. Lyon; Paris: IAC.
- VANHOVE, Claudia, 2017. *A diachronic view on suppletion. Diachronical research regarding the lexical competition between rebënok and ditja* [PhD Thesis]. University of Gent.

- VATER, Heinz, 1995. The Tense System of Polish. In: THIEROFF, R. ed., *Tense Systems in European Languages II*. Berlin/Boston: De Gruyter, Inc, pp. 153-164.
- VIELLARD, Stéphane, 2011. Problèmes de syntaxe des numéraux dans quelques langues slaves (vieux slave, vieux russe, BCMS, polonais, russe moderne), introduction. *Revue des études slaves* vol. 82, no. 4, pp. 597-603.
- VJARXOŬ, P. V., 1961. *Ličëbnik u belaruskaj move (paraŭnal'na z ruskaj i ŭkrainskaj movami)*. Minsk: Beldzjaruniversitëta imja Ul. I. Lenina.
- WHALEY, Marika L., 2000. *The evolution of the Slavic 'be(come)'-type compound future* [PhD Thesis]. The Ohio State University.
- WOODBURY, Anthony C., 2003. Defining Documentary Linguistics. *Language Documentation and Description* vol. 1, pp. 35-51.
- XOLODILOVA, Marija A., 2015. Ščetnye formy malogo količestva v russkom jazyke: zakonomernosti leksičeskogo raspredelenija. In: *80-ij jubilej Andreja Anatol'eviča Zaliznjaka*. (Online): Insitut Slavovedenie RAN. Available from: <http://inslav.ru/zalizniak80/congratulations/Kholodilova.pdf>
- ZALIZNJAK, Andrej A., 1973. O ponimanii termina "padež" v lingvističeskix opisanijax. In ZALIZNJAK, A. ed. *Problemy grammatičeskogo modelirovanija*. Moscow: Nauka, pp. 53-87.
- ZALIZNJAK, Andrej A., 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju*. Moscow: Nauka.
- ŽOLOBOV, Oleg F., 2003. K istorii malogo kvantitativa: adnumerativnye formy prilagatel'nyx i suščestvitel'nyx. *Russian Linguistics* vol. 27, pp. 117-197.
- ŽOLOBOV, Oleg F., 2007. *Drevneslavjanskije čislitel'nye kak čast' reči. Učebnoe posobie po speckursu "Sravnitel'no-istoričeskoe jazykoznanie: čislitel'nye*. Kazan: Kazan'skij Gosudarstvennyj Universitet.
- ŽOLOBOV, Oleg F. and KRYŠKO, Viktor B., 2001. *Dvojstvennoe čislo*. Moscow: Azbukovnik.

Other resources referred to in this work

ARKUŠYN, Hryhorij, 2016. *Slovnyk zaxidnopolis'kyx hovirok*. 2nd ed. Lutsk: Sxidnojevropejs'kyj nacional'nyj universytet im. Lesi Ukrajin'ky. Zaxidnopolis'kyj onomastyko-dialektolohičnyj centr.

Belarus, 2007. *National System of Geographic Names Transmission into Roman Alphabet in Belarus*. 2007, July 10, Item16(a) of the provisional agenda.

BOLDYRJEV, R., KOLMIJEC, V., LUKINOVA, T., MEL'NYČUK, O., SKLJARENKO, V., STOJANOV, I. and ŠAMOTA, A., 2006. *Etymolohičnyj slovník ukrajinskoj movy*. In: MEL'NYČUK, O. ed. Naukova dumka.

BRÜCKNER, Aleksander, 1927. *Słownik etymologiczny języka polskiego*. Kraków: Krakowska Spółka Wydawnicza.

HRINČENKO, Borys, 2010. *Slovar' ukrajin'skoj movy*. HRINČENKO, B. ed. 2 (electronic) ed. Kiev: Kievskaja Staryna.

Hrvatski Jezyčni Portal, 2006-2019. (Online): Srce; Znanje. Available from: <http://hjp.znanje.hr>

HSBM, 1981 - [ongoing]. *Histaryčny sloŭnik belaruskaj movy*. BULYKA, A. M. ed. Minsk: Belaruskaja navuka.

Institute of Russian language, 2003-2019. *Russian National Corpus*. (Online): Institute of Russian language, Russian Academy of Sciences. Available from: <http://www.ruscorpora.ru/>

KLIMČUK, Fëdar D., 2010. *Novyj Zavet. Perevod na zapadnopolesskij govor*. Minsk: Mëdžik.

MAKSYMIUK, J., 2007. Fonetyka i fonologia. In: *Zarys pisowni i gramatyki języka podlaskiego*. (Online). www.svoja.org.

National Statistical Committee of the Republic of Belarus (Belstat), 2018. *Official statistics. Social sector. Demography*. Available from: <http://www.belstat.gov.by>.

Rečnik na bǎlgarskija ezik, (ongoing). (Online): Institutăt za bǎlgarski ezik. Available from: <http://ibl.bas.bg/rbe/lang/bg/>

ŠYLJAHovič, M., 1990. Pryncypy zložinnja jityvežij (polis`kij) lytyrac`kij movy. In: *Jityveža (polis`ka) študijno-praxtyc`ka konfyrèncija*.

The Bible. New International Version (2011). (Online): Biblica Inc. Available from: <https://www.biblegateway.com/versions/New-International-Version-NIV-Bible/>

Online Latin Dictionary, 2003-2022. *Latin Dictionary*. (Online): Olivetti Media Communication. Available from: <https://www.online-latin-dictionary.com/>.

Oxford English Dictionary, 2018. *OED Online*. (Online): Oxford University Press. Available from: <http://www.oed.com>.

State Committee of the Property of the Republic of Belarus., 2007. *O vneseni izmenenij i dopolnenij v Instrukciju po transliteracii geografičeskix nazvanij Respubliki Belarus` bukvami latinskogo alfavita*. Republic of Belarus, 2007, June 11.

UNESCO, 2011. *Atlas of the World's Languages in Danger*. Available from: <http://www.unesco.org/culture/languages-atlas/index.php>.

VASMER, M., 1965. *Ètimologičeskij slovar` russkogo jazyka. (Perevod s nemeckogo s dop.)*. Moscow: Progress.

VJARENIČ, Vjačaslaŭ L., 2009. *Paleski arxiŭ. Linhvistyčnyja, etnagrafičnyja i historyčnyja materyjaly*. In: KLIMČUK, F. D., SMUŁKOWA, E. and ENGELKING, A. eds., Minsk: Zaxadnapaleskae navukova-krajaznaŭčae tavarystva Zaharoddze.

WALDENFELS, R. and MEYER, R., 2011. *ParaSol, a Corpus of Slavic and Other Languages*. (Online): University of Bern/University of Regensburg. Available from: parasol.unibe.ch.

Appendix I. List of participants

This list includes speakers mentioned in this thesis, who have allowed me to make their identity public (according to the name they wish to be known by). I classify them by their origin (or place in which the interview took place). I include the local toponym in brackets, when it differs from the official form.

Aljaksiejevičy (Sičyv), Imianin (Torokan) and Tatarja in Drahičyn

T2 Mazuka, Mikalaj A.

T3 Malevič, Mixail Ja.

T5 Mazuka, Kacjaryna I.

T6 Barysjuk, Maryja I.

T7 Marusja

T8 Misievič, Marya

T9 Hardzejčuk, Nadzeja I.

T11 Suško, Natalja A.

T12 Sakaloŭska, Maryja P.

Tor 1 Baba Ženja

Xab1 Lena Ivanoŭna

Bahdanaŭka (Bodanyŭka) in Luniniec

B2 Savič, Ivan

B5 Kavalevič, Aljaksej

B6 Kavalevič, Maryja A.

B8 Savič, Hanna V.

B9 Kavalevič, Viktor A.

B10 Iljučyk, Anton A.

B11 Meljux, Jaŭdokija P.

B12 Meljux, Scjapan I.

B13 Savič, Vasilij Mikalaievič

B14 Jarmol'čyk, Ivan K.

B15 Meljux, Hanna A.

B16 Meljux, Jaŭhènja S.

B17 Kavalevič, Vol'ha M.

B19 Kavalevič, Halina U.

B20 Savič, Vasilij Maksimavič

B21 Kavalevič, Nina A.

Haloučyci, Tolkovo (Drahičyn)

TL1 Kanavalčyk, Vera I.

TL4 Ryžuk, Maryja Ja.

TL6 Hrynvevič, Vera S.

HL1 Mancèvič, Dzina I.

HL2 Hanna

HL3 Novik, Ivan

HL4 Makarčuk, Simjaon D.

Semekhavičy and Žydča (Pinsk)

Z1 Bahdanovič, Vera S.

Z2 Haŭrylčuk, Nina

Z4 Snupok, Valjancina E.

Z7 Masejčuk, Nadzeja R.

Z8 Špimol', Vera V.

Z9 Kandracjuk, Zoja P.

Z10 Žydečkaja, Sjarafima I.

Vostraŭ and Pare (Pinsk)

P2 Maryja Uladzmiraŭna

Appendix II. Map of villages covered in this research



Figure 18 Villages surveyed in this project in the region of Brest (Belarus)